



**Universidade de
Aveiro**
2009

Departamento de Electrónica,
Telecomunicações e Informática

**Sérgio Miguel
do Vale Pires**

**Impacto do scaling da tecnologia CMOS no
desenho de circuitos digitais**



**Sérgio Miguel
do Vale Pires**

Impacto do scaling da tecnologia CMOS no desenho de circuitos digitais

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica do Doutor Ernesto Fernando Ventura Martins, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor Luís Filipe Mesquita Nero Moreira Alves, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

Aos meus pais e irmã.

o júri

presidente

Prof. Doutor Dinis Gomes de Magalhães dos Santos
Professor Catedrático de Departamento de Electrónica, Telecomunicações e
Informática da Universidade de Aveiro

Prof. Doutor Pedro Nuno Mendonça dos Santos
Instituto de Telecomunicações de Lisboa

Prof. Doutor Ernesto Fernando Ventura Martins
Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática
da Universidade de Aveiro

Prof. Doutor Luís Filipe Mesquita Nero Moreira Alves
Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática
da Universidade de Aveiro.

agradecimentos

Aos meus pais e irmã, Manuel Pires, Maria Pires e Cristiana Pires por todo o incentivo, conselhos, apoio incondicional e amizade que me deram tanto nos bons como nos maus momentos ao longo destes 23 anos e que me ensinaram a crescer.

Não me posso esquecer dos meus amigos que acompanharam a minha vida académica, pessoal e não só, em especial à Matricula de 2004 do curso que agora termino.

Um particular agradecimento aos meus orientadores, Doutor Ernesto Martins e o Doutor Luís Nero Alves por todo o suporte e conhecimentos que me transmitiram para a elaboração desta dissertação. Não quero deixar de referir a organização, competência e acompanhamento do trabalho que ia sendo realizado

Por fim quero agradecer à Universidade de Aveiro e ao Instituto de Telecomunicações - pólo Aveiro pelas excelentes condições oferecidas ao longo do meu percurso académico.

O meu muito obrigado a todos.

Sérgio Miguel do Vale Pires

palavras-chave

CMOS, scaling, razão de desenho, efeito de Miller

resumo

O inversor CMOS tem um importante papel no desenho de circuitos digitais. As várias portas/circuitos lógicos CMOS são construídas com base em inversores de referência. O inversor CMOS é otimizado de maneira a que haja um equilíbrio entre os seus tempos de propagação. Factores como o dimensionamento dos transístores, capacidades de carga, capacidade de Miller, *scaling* ou efeitos de segunda ordem podem causar desequilíbrios nos tempos de propagação.

Nesta dissertação pretende-se estudar o efeito que a capacidade de carga, a capacidade de Miller e a alteração das dimensões físicas dos transístores tem neste desequilíbrio.

keywords

CMOS, scaling, aspect ratio, Miller effect

abstract

The CMOS inverter plays an important role in digital circuits design. Several logic CMOS gates/circuits are built based on reference inverters. The CMOS inverter is optimized in a way that there is a balance between its propagation times. Factors such as transistors dimensions, load capacities, Miller effect, scaling or second order effects may cause imbalance in propagation times. This dissertation intends to study the effect that load capacity, Miller effect and changing physical dimensions of the transistors have in this imbalance.

*"A mente que se abre a uma nova ideia jamais voltará
ao seu tamanho original"*
Albert Einstein

Índice

Capítulo 1 – Introdução	1
1.1 Impacto do Scaling nas Tecnologias Actuais.....	3
1.2 Objectivos e Motivação.....	4
1.3 Metodologia	4
1.4 Trabalho original da dissertação.....	4
1.5 Estrutura da Dissertação	5
Capítulo 2 – Scaling do Transístor MOS.....	7
2.1 Modelo Quadrático (Modelo Shichman - Hodges).....	8
2.2 Efeitos de Segunda Ordem.....	9
2.2.1 Efeito de Corpo (body - effect)	9
2.2.2 Inversão Fraca (Weak Inversion Mode).....	10
2.2.3 Modulação do Comprimento de Canal.....	11
2.2.4 Saturação da Velocidade de Deriva.....	12
2.2.5 Degradação da Mobilidade dos Portadores de Carga	13
2.2.6 DIBL (Drain-Induced Barrier Lowering).....	14
2.2.7 Punch – Through	14
2.2.8 CMOS Latchup	15
2.3 Capacidades do MOSFET.....	16
2.3.1 Capacidades do Canal.....	17
2.3.2 Capacidades Estruturais (Capacidades de Overlap)	18
2.3.3 Capacidade das Difusões	19
2.3.4 Modelo de Capacidades do Dispositivo.....	20
2.3.5 Comportamento das Ligações	21
2.4 Scaling.....	22
2.4.1 Scaling Completo (ou Scaling de Campo Eléctrico constante).....	23
2.4.2 Scaling por Tensões Fixas.....	23
2.4.3 Scaling Geral	24
2.5 Impacto do Scaling nos Dispositivos e Resolução de Alguns Problemas	25
2.6 Sumário	26
Capítulo 3 – Desenho de Portas Lógicas CMOS.....	27
3.1 Portas CMOS – Conceito.....	28

3.2	Porta NAND.....	28
3.2.1	Característica Estática da NAND2.....	29
3.2.2	Tempos de Propagação da NAND2.....	30
3.3	Porta NOR.....	31
3.3.1	Característica estática da NOR2.....	32
3.3.2	Tempos de Propagação da NOR2.....	32
3.4	Comparação entre as portas NAND e NOR.....	33
3.5	Efeito de Miller.....	34
3.5.1	Modelo de carga simples.....	34
3.5.2	Modelo Incluindo a Capacidade de Miller do Inversor	35
3.6	Sumário	39
Capítulo 4 – Resultados		41
4.1	Tecnologia CMOS Utilizada	42
4.2	Método de Desenvolvimento	42
4.3	Conjuntos de Teste	44
4.4	Resultados ao Nível do Layout.....	46
4.4.1	Caracterização de um Inversor Estático	46
4.4.2	Estudo do Comportamento das Portas Lógicas com uma Carga Linear	47
4.4.3	Comportamento das Portas Lógicas Tendo um Inversor como Carga	49
4.5	Resultados ao Nível da Simulação Esquemática	53
4.6	Sumário	57
Capítulo 5 – Conclusões		59
5.1	Linhas de investigação futuras.....	60
Referências.....		61
Anexos 1		63

LISTA de Figuras

<i>Figura 2.1</i> – Vista de um corte de um MOSFET tipo N	8
<i>Figura 2.2</i> – Corrente na região sublimiar	10
<i>Figura 2.3</i> – Curva da modulação do comprimento de canal.....	11
<i>Figura 2.4</i> – Saturação por velocidade de deriva	12
<i>Figura 2.5</i> – Curvas característica num dispositivo com problemas de <i>punch-through</i>	14
<i>Figura 2.6</i> – CMOS <i>Latch-up</i> e respectivo circuito equivalente	15
<i>Figura 2.7</i> – Capacidades do MOSFET: corte vertical	16
<i>Figura 2.8</i> – Capacidades da porta para o canal e região de funcionamento do MOSFET	17
<i>Figura 2.9</i> – Capacidade de overlap do MOSFET, vista de cima (esquerda) e vista de corte (direita).....	18
<i>Figura 2.10</i> – Vista em detalhe da junção da fonte	19
<i>Figura 2.11</i> – Modelo das capacidades parasitas do MOSFET e respectivas equações	20
<i>Figura 2.12</i> – Estrutura de uma ligação.....	21
<i>Figura 2.13</i> – Exemplo de seis linhas de interligação em três níveis diferentes	21
<i>Figura 2.14</i> – Influência do C_{fringe} nas ligações.....	22
<i>Figura 2.15</i> – Evolução da diminuição do comprimento do canal	22
<i>Figura 3.1</i> – Estrutura de uma porta lógica CMOS	28
<i>Figura 3.2</i> – Esquema eléctrico, símbolo e função lógica de uma NAND de duas entradas	29
<i>Figura 3.3</i> – Várias VTC's de uma NAND2	29
<i>Figura 3.4</i> – Esquema eléctrico, símbolo e função lógica de uma NOR de duas entradas.....	31
<i>Figura 3.5</i> – Diferentes VTC's de uma NOR2.....	32
<i>Figura 3.6</i> – Esquema básico de um inversor CMOS, com a respectiva capacidade de Miller.....	34
<i>Figura 3.7</i> – Inversor CMOS tendo outro inversor CMOS como carga	37
<i>Figura 4.1</i> – Histograma sobre o funcionamento do Cadence como ferramenta de desenvolvimento	43
<i>Figura 4.2</i> – Ambiente de simulação com carga capacitiva	44
<i>Figura 4.3</i> – Ambiente de simulação usando um inversor como carga	44
<i>Figura 4.4</i> – Ambiente de simulação com um inversor ideal como carga	45
<i>Figura 4.5</i> – Ambiente de simulação tendo como carga uma porta lógica	46
<i>Figura 4.6</i> – Esquema eléctrico de uma NAND2, NOR2, NAND3 e NOR3	47

Figura 4.7 – Resposta das várias portas lógicas à variação de cargas capacitivas lineares.....	48
Figura 4.8 – Resposta das portas lógicas a um inversor de carga	49
Figura 4.9 – Desequilíbrio da NAND2 em função de do tamanho do inversor ideal	50
Figura 4.10 – Desequilíbrio em função do tamanho dos inversores ideais usando capacidade de Miller.....	51
Figura 4.11 –Desequilíbrio em função do tamanho dos inversores para dois valores de VM diferentes tendo como driver um NAND2_8X	52
Figura 4.12 – Desequilíbrio em função do tamanho da porta driver (esquerda), Desequilíbrio em função de uma NAND como carga (meio), Desequilíbrio em função de uma NOR como carga (direita).....	52
Figura 4.13 – Capacidade de Entrada equivalente	54
Figura 4.14 – Dependência entre os tempos de propagação e o tempo de subida (esquerda – esquema de simulação da figura 4.2, direita esquema de simulação da figura 4.3)	54
Figura 4.15 – Desequilíbrio em relação às condições de carga (esquerda – esquema de simulação da figura 4.2, direita esquema de simulação da figura 4.3).....	55
Figura 4.16 – Desequilíbrio em função das razões de desenho (esquerda – inversor de driver, direita – inversor de carga).....	56

Lista de Tabelas

Tabela 2.1 – Distribuição das capacidades do canal do MOSFET nas diferentes regiões..... 17

Tabela 2.2 – Influência do scaling nos diferentes parâmetros do MOSFET..... 25

Lista de acrónimos

ADE	<i>Analog Design Environment</i>
AMS	<i>Austria Microsystem</i>
BOX	<i>Buried Oxide</i>
BJT	<i>Bipolar Junction Transistor</i>
CMOS	<i>Complementary Metal-Oxide Semiconductor</i>
CAD	<i>Computer-Aided Design</i>
DGSOI	<i>Double Gate Silicon-on-Insulator</i>
DIBL	<i>Drain-Induced Barrier Lowering</i>
DRC	<i>Design Rule Check</i>
DTMOS	<i>Dynamic Threshold Metal-Oxide Semiconductor</i>
FET	<i>Field-Effect Transistor</i>
FILOX	<i>Fillet Local Oxidation</i>
GAA	<i>Gate All Around</i>
LVS	<i>Layout Vs. Schematic</i>
MOSFET	<i>Metal-Oxide Semiconductor Field-Effect Transistor</i>
PDN	<i>Pull-Down Network</i>
PUN	<i>Pull-Up Network</i>
RCX	<i>Resistance/Capacitance and Inductance Extraction</i>
SOI	<i>Silicon-on-Insulator</i>
VLSI	<i>Very Large Scale Integration</i>
VTC	<i>Voltage Transfer Characteristic</i>

Lista de Símbolos

β	Razão geométrica de desenho
C_{DB}	Capacidade dreno-substrato
C_{fringe}	Capacidade lateral
C_G	Capacidade da porta
C_{GC}	Capacidade da porta para o canal
C_{GCB}	Capacidade da porta para o substrato
C_{GCS}	Capacidade da porta para a fonte
C_{GDO}	Capacidade de sobreposição entre a porta e o dreno
C_{GSO}	Capacidade de sobreposição entre a porta e a fonte
C_{inter}	Capacidade das interligações
C_j	Capacidade da junção da base
C_{jsw}	Capacidade da junção lateral
C_L	Capacidade de carga
C_M	Capacidade de Miller
C_{ML}	Capacidade de Miller da carga
C_o	Capacidade por unidade de área
C_{ox}	Capacidade porta por unidade de área
C_{pp}	Capacidade de placas paralelas
C_{SB}	Capacidade da fonte para o substrato
C_{wire}	Capacidade ligações
δ	Desequilíbrio relativo dos tempos de propagação
ξ_c	Campo eléctrico crítico
ξ	Campo eléctrico
ϵ_s	Permitividade do silício
FO	<i>Fan-out</i>
GND	Tensão de referência (massa)

I_C	Corrente de colector
I_{DN}	Corrente de dreno de um NMOS
I_{DP}	Corrente de dreno de um PMOS
I_{DS}	Corrente de dreno
I_{DS0}	Corrente de dreno quando $V_{GS} = V_{th}$
I_o	Corrente de saída
Φ_F	Potencial de Fermi
k	Transcondutância de um MOSFET
k_B	Constante de Boltzman
λ	Factor de modulação de comprimento de canal
L	Comprimento de canal
L_{eff}	Comprimento de canal efectivo de um MOSFET
N_A	Concentração de impurezas
q	Carga do electrão
R_N	Resistência equivalente de um MOSFET tipo N
R_P	Resistência equivalente de um MOSFET tipo P
θ	Coeficiente de degradação da mobilidade dos portadores
T	Temperatura
t_{ox}	Espessura do óxido
t_p	Tempo de propagação
t_{pHL}	Tempo de propagação de <i>high-to-low</i>
t_{pLH}	Tempo de propagação de <i>low-to-high</i>
μ	Mobilidade dos portadores de carga
V_a	Tensão de saída do inversor de carga
V_{be}	Tensão entre a base e o emissor
v_d	Velocidade de deriva
V_{DD}	Tensão de alimentação positiva

V_{DS}	Tensão entre o dreno e a fonte
V_{GD}	Tensão entre a porta e o dreno
V_{GS}	Tensão entre a porta e a fonte
V_i	Tensão de entrada
V_M	Tensão de comutação de um inversor
V_o	Tensão de saída
V_{SB}	Tensão entre a fonte e o substrato
V_{TH}	Tensão de limiar
V_{TH0}	Tensão de limiar para $V_{SB} = 0$
W	Largura do canal de um transistor
x_d	Difusão lateral
X_j	Profundidade da junção
γ	Factor de corpo

Capítulo 1 – Introdução

A longa história da microelectrónica foi construída num curto espaço de tempo, e teve um grande impacto na nossa sociedade. Citado na edição de Janeiro de 1998 da revista “*Proceedings of the IEEE*”, o transistor é considerado como “a invenção da engenharia electrónica mais revolucionária do século XX, cujo impacto é sentido a todo o momento, em todo o lugar na era da informação”.

A constante redução do tamanho dos transístores impulsionada por avanços significativos na indústria dos semicondutores fizeram aumentar muito a criação de sistemas VLSI (*Very Large Scale Integration*).

Em 1952, numa conferência proferida no encontro anual de componentes, ocorrida nos EUA, o cientista inglês Geoffrey W.A. Dumer, considerando o advento do transistor, já previa o aparecimento, num futuro não muito distante, de aparelhos electrónicos, formados por conjuntos de componentes isentos de interligações por fios e que estariam montados e inseridos num bloco sólido. Esta previsão tecnológica não tardou acontecer, quando dois engenheiros americanos, Jack Kilby, da *Texas Instrument*, e Robert Noyce, da *Fairchild Semiconductor*, trabalhando independentemente concebem, em 1959 os primeiros circuitos integrados [12,13]. Kilby fê-lo numa placa de germânio enquanto no conceito proposto pelo Noyce foi usado como semicondutor base o silício.

A produção em massa de circuitos integrados começou por volta de 1960. Entretanto, eram ainda considerados semicondutores caros apresentando certas deficiências operacionais devido às características dos sistemas usados para acoplar os diversos transístores. Assim, a tecnologia CMOS (*Complementary Metal-Oxide Semiconductor*)[1], advinda das pesquisas com o transístor de efeito de campo, MOSFET, é rapidamente desenvolvida e utilizada na elaboração de semicondutores com elevada capacidade de integração, podendo ser incorporados numa área extremamente reduzida um maior número de portas.

Gordon Earle Moore co-fundador da *Intel Corporation*, autor da Lei de Moore[14] (publicada na *Electronics Magazine* num artigo de 19 de Abril de 1965), constatou que a cada 18 meses a capacidade de processamento dos computadores dobra, enquanto os custos permanecem constantes.

Contudo, para que a lei de Moore fosse válida era imprescindível que as reduções físicas dos transístores fossem reduzidas a fim de na mesma área de silício fosse possível colocar mais componentes (aumento da área de integração). Robert Dennard foi o primeiro a introduzir a discussão sobre este tema, que ficaria conhecido por *scaling*[15], afirmando que se todas as dimensões físicas e tensão de funcionamento dos transístores fossem reduzidas por um mesmo factor, os dispositivos resultantes não apresentariam alteração das características eléctricas e funcionais. De facto, este modelo garantiu o fabrico de transístores mais pequenos e com desempenho ao nível de velocidade, consumo de potência e custos de produção menores.

A Lei de Moore foi seguida durante quatro décadas, mas a diminuição das dimensões dos transístores levou ao aumento das correntes de fuga, o que, por sua vez, levou a que fossem adoptadas outras técnicas, já que a gestão das correntes de fuga é crucial para a operação confiável dos *chips* [16]. A nova tecnologia “*high-k*”, foi a solução desenvolvida pela *Intel* para os seus transístores de 45nm. A principal mudança é a utilização de um metal chamada háfnio que vai substituir o dióxido de silício (SiO_2) usado na construção da porta do MOSFET. Com esta tecnologia conseguiu-se uma melhoria de cerca de 20% na velocidade de comutação bem como uma redução das perdas pela porta do transístor em 10% [16].

A *Intel* também apresentou uma SRAM funcional com transístores de 32nm que contém mais de 1,9 milhões de transístores *high-k metal gate*. O lançamento dos transístores de 32nm está previsto para 2009.

Segundo a ITRS (International Technology Roadmap for Semiconductors) as tecnologias seguintes previstas serão a de 22nm e tem lançamento agendado para 2011-2012, 16nm (apontado para 2018) e 11nm (esperada para 2022).

1.1 Impacto do Scaling nas Tecnologias Actuais

O *scaling* da tecnologia trouxe vários desafios aos desenhadore de circuitos. Se por um lado se obtiveram grandes vantagens como a maior densidade dos circuitos e a maior velocidade das transições lógicas, por outro lado também houve perdas devido ao aumento da influência dos efeitos secundários nos circuitos.

Os transístores de canal curto são muito afectados por efeitos secundários como a saturação de velocidade de deriva e a degradação da mobilidade dos portadores de carga. O aumento das correntes de fuga é um efeito secundário que trouxe bastantes implicações nos circuitos uma vez que a potência estática consumida aumenta bastante e além disso podem provocar a alteração dos estado lógico da porta.

Devido a estes factos, as técnicas tradicionais de desenho tornaram-se algo desadequadas, uma vez que não entram em consideração com estes efeitos provocados pelos dispositivos de canal curto, e por conseguinte deixaram de permitir o desenho de circuitos digitais optimizados.

O trabalho desta dissertação incide sobre o impacto do scaling nas técnicas tradicionais e o estudo e criação de um modelo matemático que permita identificar a influência do efeito de *Miller* presente nas portas lógicas.

1.2 Objectivos e Motivação

A evolução tecnológica descrita anteriormente bem como a contínua redução das dimensões físicas fez com que apenas a discussão sobre os transístores de canal curto fizesse sentido. Contudo o aparecimento destes transístores fez com que o modelo de desenho usado até então deixasse de ser válido uma vez que os dispositivos não cumpriam com os requisitos propostos.

É neste sentido que surge este trabalho de dissertação, isto é, um estudo dos efeitos causados pelo *scaling* na optimização das portas lógicas bem como o desequilíbrio entre os tempos de propagação provocado nessas mesmas portas pela capacidade de *Miller* e pelas alterações de VM.

1.3 Metodologia

O objectivo deste trabalho é avaliar o impacto que o *scaling* provoca no desequilíbrio das portas lógicas usando uma tecnologia de canal curto. O plano proposto consiste inicialmente no estudo e caracterização de algumas portas lógicas e o estudo do efeito de *Miller* usando para realização o *design-kit* promovido pela AMS (*Austria Microsystems*), ao abrigo do protocolo celebrado entre a Universidade de Aveiro e o *Europractice*.

Os vários ambientes de simulação criados bem como os layouts foram desenvolvidos através do ambiente integrado no *Cadence DFII (Design Framework II)*. A caracterização das várias portas bem como do efeito de *Miller* foi feita também em ambiente *Cadence*, utilizando neste caso o simulador *Spectre* juntamente com a interface com o utilizador ADE (*Analog Design Environment*).

1.4 Trabalho original da dissertação

No desenvolvimento desta dissertação foi escrito um artigo com o título “*A Study on the Propagation Times of Loaded CMOS Inverters*” [25], submetido para a conferência Iscas 2010 (www.iscas2010.org), que aborda a problemática da capacidade de *Miller* e da razão de desenho no equilíbrio dos tempos de propagação nos circuitos digitais.

1.5 Estrutura da Dissertação

Esta dissertação divide-se em cinco capítulos. No primeiro capítulo faz-se uma introdução sobre o tema do trabalho onde se faz uma breve descrição histórica sobre a evolução do transistor e processos litográficos, a importância da lei de Moore na evolução dos circuitos integrados e as teorias de scaling propostas por Dennard. Descreve-se também as perspectivas futuras da tecnologia e expõem-se a motivação e os objectivos.

No segundo capítulo é apresentado o modelo de *Shichman - Hodges* dos MOSFET's, são também explicados os efeitos secundários a que os dispositivos estão sujeitos. O estudo destes efeitos é essencial para a compreensão do comportamento dos circuitos principalmente quando se fala de transístores de canal curto, já que é devido a eles que os MOSFET's deixaram de ter um comportamento totalmente conhecido. Posteriormente são dadas a conhecer vários modelos de *scaling*, bem como o efeito que estas causam na física dos dispositivos.

No capítulo 3 é apresentado o conceito dos circuitos CMOS estáticos. Começa por se mostrar como são criadas as redes PUN (*Pull-Up Network*) e PDN (*Pull-Down Network*). De seguida são dadas a conhecer portas NAND e NOR, estudando as suas características mais importantes, tanto ao nível estático como a nível dinâmico. É também apresentada a teoria tradicional de desenhos de circuitos combinatórios. Depois faz-se uma abordagem sobre a capacidade de *Miller* tanto na porta *driver* como na porta de carga e apresenta-se um modelo matemático que justifica a influência causada nos tempos de propagação por esta capacidade. O fim do capítulo é dedicado ao estudo sobre a influência que a variação da tensão de threshold pode causar nos tempos de propagação.

O capítulo 4 é referente aos resultados do trabalho. Neste capítulo é explicado todo o trabalho prático desenvolvido, e a metodologia de desenvolvimento e estratégias seguidas ao longo do trabalho. São também descritas todas as simulações e interpretados os resultados obtidos.

Por fim o capítulo 5 faz um resumo de toda a dissertação, apresentando as conclusões mais importantes retiradas do trabalho. São também sugeridas algumas linhas de investigação futura de forma a dar continuação ao estudo efectuado.

Capítulo 2 – Scaling do Transístor MOS

O MOSFET (*Metal-Oxid Semiconductor Field Effect Transistor*) surgiu em 1925, proposto por *Julius Edgar Lilienfeld*, muito antes do BJT (*Bipolar Junction Transistor*), mas a falta de tecnologia e materiais adequados atrasou o seu uso prático para a década de 60.

Em 1963 *Frank Wanlass* e *Chih-Tang Sah* [2], apresentaram uma tecnologia revolucionária que usava PMOS e NMOS simultaneamente, ficando conhecida por CMOS (*Complementary Metal Oxide Semiconductor*), contrariando as tecnologias até então usadas, que utilizavam apenas NMOS. O aumento da escala integração, o bom desempenho nas transições de estado lógico, o baixo consumo estático entre outras vantagens, fizeram do CMOS a tecnologia dominante até à década de 90, chegando a compor mais de 95% dos circuitos integrados digitais.

Neste capítulo pretende-se estudar alguns efeitos que se começaram a notar mais com o aparecimento de transístores de canal curto. Em uma primeira abordagem, vai ser apresentado o modelo quadrático do MOSFET, seguindo-se o estudo dos efeitos de segunda ordem, capacidades parasitas, o comportamento das ligações e por fim o conceito de *scaling*.

2.1 Modelo Quadrático (Modelo Shichman - Hodges)

O MOSFET (*Metal-Oxide Semiconductor Field-Effect Transistor*) é o componente base de qualquer circuito digital actual. Devido ao seu fácil fabrico, o seu custo de produção é reduzido. As principais características do MOSFET são o baixo consumo de potência estática e pela reduzida área de silício ocupada por cada transistor [18].

A figura 2.1 representa uma vista em corte de um transistor do tipo N (cujo o substrato é de tipo P), existe também um transistor de tipo P que se identifica devido ao seu substrato ser de tipo N. Nesta são apresentados os quatro terminais do transistor MOS que são S, G, D e B, que se designam respectivamente por fonte (*Source*), porta (*Gate*), dreno (*Drain*) e substrato (*Body*).

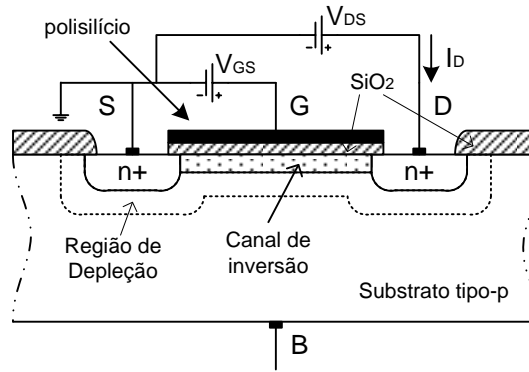


Figura 2.1 – Vista de um corte de um MOSFET tipo N

Ao longo dos anos, os circuitos foram estudados com base nas equações do modelo quadrático do MOSFET [2, 3] que é o modelo mais simples e que permite o cálculo manual além de dar uma primeira aproximação do comportamento estático do transistor. Para transístores de canal N estas equações são:

$$I_{DS} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} [2(V_{GS} - V_{th})V_{DS} - V_{DS}^2], \quad (2.1)$$

Quando $V_{GS} > V_{th}$ e $V_{GD} > V_{th}$, zona linear e

$$I_{DS} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} (V_{GS} - V_{th})^2, \quad (2.1)$$

quando $V_{GS} > V_{th}$ e $V_{GD} < V_{th}$, zona de saturação [18],

$I_{DS} \rightarrow$ corrente entre o dreno e a fonte

$\mu_n \rightarrow$ mobilidade dos portadores de carga

$C_{ox} \rightarrow$ capacidade do óxido por unidade de área

$W \rightarrow$ largura do canal

$L \rightarrow$ comprimento do canal

$V_{GS} \rightarrow$ tensão aplicada entre a porta e a fonte

$V_{th} \rightarrow$ tensão de limiar

$V_{DS} \rightarrow$ tensão entre o dreno e a fonte.

2.2 Efeitos de Segunda Ordem

O modelo quadrático oferecia uma boa aproximação do real comportamento do MOSFET, mas com a diminuição do comprimento de canal dos transístores, alguns efeitos que eram desprezados na equações, começaram a fazer-se notar cada vez mais, o que aumenta a importância do estudo destes efeitos, denominados efeitos de segunda ordem.

2.2.1 Efeito de Corpo (body - effect)

Idealmente o terminal da fonte deve estar ligado ao terminal do substrato, o que implica uma tensão fonte - substrato igual a zero. Porém na prática, o substrato é ligado ao terminal de menor tensão para os NMOS e de maior tensão no caso dos PMOS, isto porque torna mais simples e prático ter um substrato e um poço comum para todo o chip, podendo a fonte estar ligada a um terminal de potencial diferente gerando assim uma tensão de fonte - substrato (V_{SB}) [19] diferente de zero o que faz variar o comprimento da região de depleção. Esta variação altera a tensão de threshold (V_{th}) segundo a equação:

$$V_{th} = V_{th0} + \gamma \sqrt{|2\Phi_F| + V_{SB}} - \sqrt{|2\Phi_F|} \quad (2.3)$$

Onde V_{th0} é a tensão de threshold quando V_{SB} é 0, Φ_F é um parâmetro físico denominado o potencial de Fermi, tipicamente 0.3V para um substrato tipo P e γ é um parâmetro do

processo de fabrico, conhecido como “factor de corpo”, o qual por sua vez é determinado através de:

$$\gamma = \frac{\sqrt{2qN_A\epsilon_s}}{C_{ox}} \quad (2.4)$$

Em que q é a carga do electrão ($1.6 \times 10^{-19} \text{C}$), N_A é a concentração de impurezas de tipo-p no substrato e ϵ_s é a permissividade do silício ($1.04 \times 10^{-12} \text{F/cm}$), o valor de γ é tipicamente $0.4 \text{V}^{1/2}$ [3].

De referir que o scaling da tecnologia é acompanhado por valores diferentes de N_A , fazendo variar por sua vez γ .

2.2.2 Inversão Fraca (Weak Inversion Mode)

Olhando de uma forma atenta para a figura 2.2:

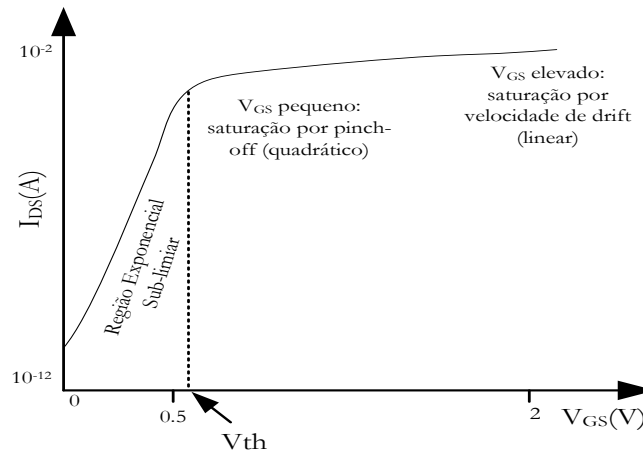


Figura 2.2 – Corrente na região sublimiar

É possível observar que a corrente quando $V_{GS} < V_{th}$ não é zero, o que significa que o transistor já se encontra em condução para valores abaixo da tensão de limiar. A este efeito é dado o nome de condução na região de sublimiar.

A corrente de sublimiar é uma função exponencial de V_{GS} e V_{DS} sendo dada pela expressão:

$$I_{DS} = I_{DS0} e^{\frac{V_{GS}-V_{th}}{nV_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}} \right) \quad (2.5)$$

Onde I_{DS0} é igual à corrente quando $V_{GS} = V_{th}$, $V_T = K_B \frac{T}{q}$, T é a temperatura, K_B é a constante de Boltzman e “n” é um parâmetro empírico, tipicamente com um valor de cerca de 1.5 [18].

2.2.3 Modulação do Comprimento de Canal

A equação (2.2) descreve uma corrente entre os terminais dreno e fonte como sendo constante e independente da tensão aplicada a esses terminais. De facto trata-se de uma simplificação que por vezes se afasta bastante da realidade, dado que o comprimento do canal é modulado pela tensão dreno - fonte. Aumentando o V_{DS} na região de saturação, ocorre um aumento da região de depleção na zona do dreno, reduzindo assim o comprimento efectivo do canal [19].

Como se observa na equação (2.2) a corrente I_{DS} aumenta com a diminuição do L , portanto uma melhor aproximação da corrente na zona de saturação será

$$I_{DS} = \frac{\mu_n C_{ox} W}{2 L} (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \quad (2.6)$$

Em que λ é um parâmetro empírico conhecido como coeficiente de modulação do comprimento de canal. [3]

Observando as curvas características I-V, na figura 2.3, nota-se que na zona de saturação o parâmetro $(1 + \lambda V_{DS})$ faz-se notar uma vez que a recta característica na região de saturação aparece agora inclinada:

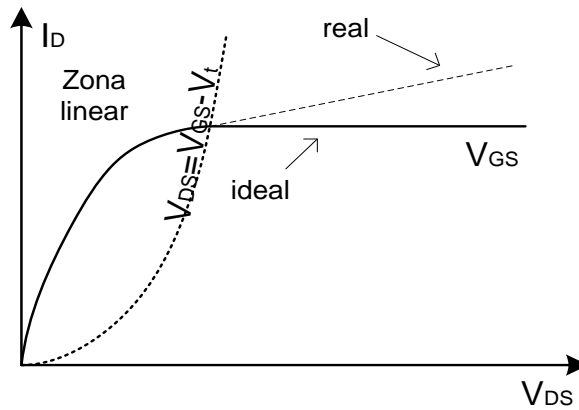


Figura 2.3 – Curva da modulação do comprimento de canal

2.2.4 Saturação da Velocidade de Deriva

O comportamento dos transístores de canal curto é bastante diferente do comportamento já conhecido dos transístores de canal longo. O principal responsável por esta diferença é o efeito de saturação por velocidade de deriva dos portadores no canal.

A equação (2.7) mostra que a velocidade dos portadores é proporcional ao campo eléctrico $|\xi|$, isto é a mobilidade dos portadores é constante. De referir que o modelo quadrático de Shichman - Hodges [2,3] apresentado anteriormente baseia-se na dependência linear entre a velocidade de deriva e campo eléctrico.

$$v_n = -\mu\xi(x) \quad (2.7)$$

Contudo um forte campo eléctrico horizontal faz com que este modelo matemático linear não seja seguido. O que acontece é que a velocidade aumenta linearmente com o efeito do campo eléctrico até um ponto crítico $|\xi_c|$, a partir do qual a velocidade tende para um valor constante, como é visível na figura 2.4. Isto deve-se ao efeito de *scattering* que é resultado das colisões sofridas pelos portadores.

Este efeito tem um impacto profundo na operação do transístor, e a equação da velocidade pode ser reescrita da seguinte forma:

$$\begin{cases} v = \frac{\mu_n \xi}{1 + \frac{\xi}{\xi_c}} & \text{para } \xi < \xi_c \\ v_{sat} = \mu_n \xi_c & \text{para } \xi \geq \xi_c \end{cases} \quad (2.8)$$

Dando origem á figura 2.4

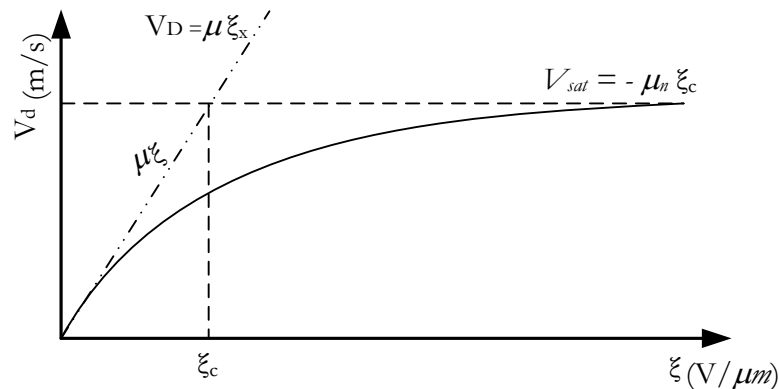


Figura 2.4 – Saturação por velocidade de deriva

Este efeito provoca grandes mudanças no comportamento do dispositivo tais como a menor corrente I_{DS} para os mesmos valores de V_{GS} ; a dependência linear que passa a existir entre I_{DS} e V_{GS} na saturação e o aparecimento de dois pontos de saturação, um por *pinch-off* e outro por saturação da velocidade de deriva. Assim a expressão 2.9 indica a dependência de I_{DS} com a velocidade:

$$I_{DS} = K_n \left[(V_{GS} - V_{th})V_{DS} - \frac{V_{DS}^2}{2} \right] \frac{1}{1 + \left(\frac{V_{DS}}{\xi_{cL}} \right)} \quad (2.9)$$

2.2.5 Degradação da Mobilidade dos Portadores de Carga

A degradação da mobilidade dos portadores de carga é um efeito de segunda ordem que afecta em especial os transístores de canal curto. Tal como o próprio nome indica, a mobilidade dos portadores de carga é a capacidade que estes têm para se deslocar num certo material, sendo determinada pela razão entre a velocidade de deriva dos portadores e o campo eléctrico longitudinal no canal.

Esta mobilidade varia consoante os portadores de carga utilizados. É sabido que no silício $\mu_n = 3\mu_p$ e no AsGa $\mu_n = 30\mu_p$ sendo μ_n e μ_p a mobilidade dos electrões e das lacunas respectivamente. Esta é a principal razão para os MOSFET de tipo N terem maior facilidade de produzir corrente [3].

A degradação de μ_n ocorre principalmente devido ao efeito do campo eléctrico vertical (transversal) no canal de inversão, já que este acelera os electrões fazendo-os chocar com a interface do canal - óxido, provocando assim uma dispersão entre a interface do substrato e do óxido da porta. A mobilidade diminui também com o aumento da temperatura e com o aumento de concentração de impurezas [3].

Um modelo matemático simples que descreve o efeito do campo eléctrico vertical sobre a mobilidade é:

$$\mu_n = \frac{\mu_0}{1 + \theta(V_{GS} - V_{th})} \quad (2.10)$$

em que μ_n é a mobilidade efectiva dos portadores e θ é o coeficiente de degradação da mobilidade pelo campo eléctrico transversal, sendo este um valor empírico.

2.2.6 DIBL (Drain-Induced Barrier Lowering)

Com a diminuição do tamanho dos transístores, algumas características dos dispositivos, começam fazer-se notar de forma mais efectiva. O estudo da região de depleção em torno do dreno e da fonte torna-se relevante já que faz variar de forma significativa a tensão de limiar V_{th} . É neste contexto que surge o termo DIBL (*Drain-Induced Barrier Lowering*), que é caracterizado como o efeito que a tensão V_{DS} tem sobre a região de depleção na zona do dreno e, consequentemente, na tensão V_{GS} de inversão do canal. O aumento de V_{DS} tende a aumentar a região de depleção reduzindo o comprimento efectivo do canal e aumentando assim a corrente I_{DS} [3, 18]. O efeito é semelhante a uma redução de V_{th} .

Analiticamente é muito complicado encontrar uma relação matemática para este efeito, mas uma aproximação razoável pode ser descrito em (2.11):

$$V_{th} = V_{th0} - \sigma V_{DS} \quad (2.11)$$

que como foi explicado anteriormente descreve uma variação na tensão de limiar, quando V_{DS} também varia, podendo o valor do DIBL variar entre 20 e 150 mV por cada volt de variação na tensão de V_{DS} .

2.2.7 Punch – Through

Tal como no DIBL, este efeito faz-se notar com o aumento do potencial electrostático entre o dreno e a fonte. Neste caso, o continuo aumento da tensão do dreno obriga ao aumento da região de depleção nessa zona de tal forma que fica próxima da região de depleção da fonte formando num caso extremo, uma região de depleção única, deixando de existir assim inversão do canal.

Devido a este facto, a porta deixa de controlar o dispositivo, havendo um fluxo de corrente, independente da tensão aplicada na

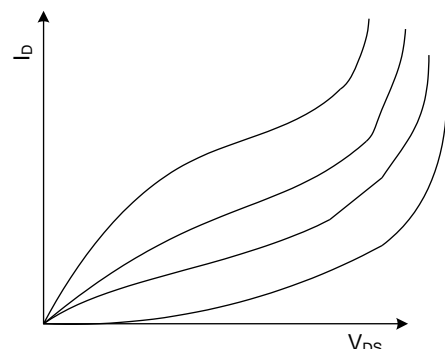


Figura 2.5 - Curvas característica num dispositivo com problemas de *punch-through*

porta, porque os portadores “perfuram” o canal de uma região, sendo este efeito conhecido como *punch – through* [18] e define o limite superior para a tensão V_{DS} que pode ser aplicado ao dispositivo.

2.2.8 CMOS Latchup

Devido ao *layout* em circuitos integrados, a tecnologia CMOS apresenta um problema, conhecido por *latch-up*, que pode ser descrito como o fenómeno de condução parasita que pode ocorrer num *chip*, e assemelha-se á operação de um tiristor. Este problema deve-se á presença de estruturas n-p-n-p.

O *latch-up* pode desencadear um curto-circuito entre V_{DD} e a massa, podendo em casos extremos levar á auto-destruição do *chip*.

Para ilustrar o *latch-up* recorre-se á figura 2.6, onde são mostradas as conexões entre o substrato e a massa, e entre o poço p e o V_{DD} . É possível notar na figura, dois BJT's (*Bipolar Junction Transistor*), e as resistências R_{psub} e R_{well} .

Considerando que ocorre um fluxo de carga na região n^+ (fonte do NMOS), dando origem a uma corrente de emissor no NPN, que faz aparecer uma queda de tensão da base para o emissor de NPN, $V_{BE(NPN)}$.

A queda de tensão $V_{BE(NPN)}$ pode aumentar através de R_{psub} e levar o transistor bipolar NPN a um regime de condução, verificando-se uma corrente de colector. A $I_{c(NPN)}$ força uma corrente através de R_{well} , que está associada á queda de tensão $V_{EB(PNP)}$, esta queda pode tornar-se grande o suficiente para colocar o PNP em condução e tem-se assim uma corrente de emissor, $I_{E(PNP)}$ e por conseguinte uma corrente de colector $I_{C(PNP)}$. Esta corrente vai reforçar a queda de tensão $V_{BE(NPN)}$. Tendo assim um efeito de realimentação positiva, uma vez que o NPN vai conduzir ainda mais, repetindo-se assim o processo.

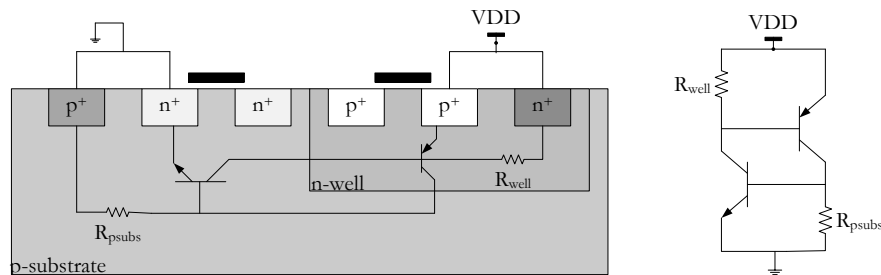


Figura 2.6 - CMOS *Latch-up* e respectivo circuito equivalente

Este efeito de *latch-up* é nas tecnologias vigentes quase insignificante uma vez que há formas de o evitar. Algumas medidas para minimizar os efeitos de *latch-up* são:

- Distanciar as regiões n^+ (MOSFET canal N) e p^+ (MOSFET canal P) para assim diminuir o ganho (β) dos BJT's
- Introduzir vários contactos p^+ para o substrato, próximos entre si, para diminuir R_{psub}
- Introduzir vários contactos n^+ para o poço n, próximos entre si, para diminuir o R_{well} .

2.3 Capacidades do MOSFET

Numa perspectiva de comportamento dinâmico, uma profunda percepção da natureza e do comportamento das capacidades intrínsecas do dispositivo é fundamental para o desenho de circuitos integrados de alta qualidade, uma vez que estas capacidades são as principais responsáveis pelos atrasos nas portas lógicas. A figura 2.7 representa um corte vertical num transístor NMOS, onde é possível verificar a existência de um conjunto de capacidades parasitas que podem ser agrupadas em três tipos de acordo com a sua natureza: a própria estrutura dos transístores (C_{GSO} , C_{GDO}), a carga dos portadores induzida no canal (C_{GS} , C_{GD} , C_{GB}) e a região de depleção formada entre o substrato com a fonte e o dreno (C_{SB} , C_{DB}). De notar que, à excepção das capacidades estruturais todas as restantes capacidades intrínsecas do transístor são não lineares dificultando muito a análise de circuitos com estes dispositivos.

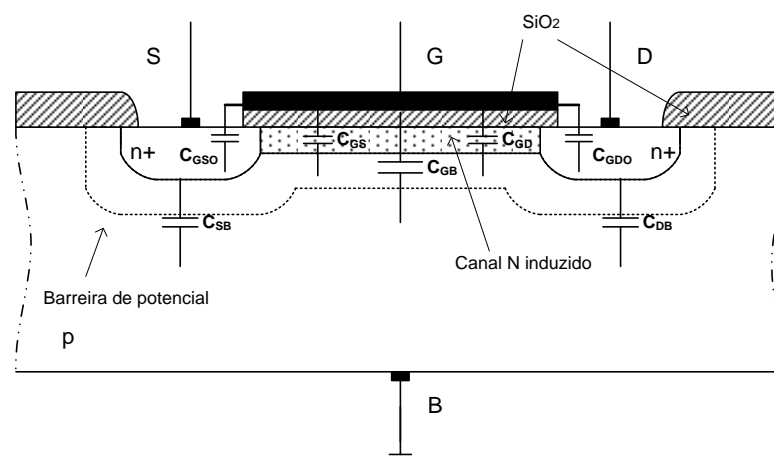


Figura 2.7 – Capacidades do MOSFET: corte vertical

2.3.1 Capacidades do Canal

As capacidades da porta para o canal (C_{GC}) podem aparecer de várias formas, dependendo da tensão aplicada a cada terminal do MOSFET, fazendo variar assim a sua região de funcionamento. Por conseguinte podem-se dividir estas capacidades em três:

- $C_{GS} \rightarrow$ Capacidade da porta para a fonte
- $C_{GD} \rightarrow$ Capacidade da porta para o dreno
- $C_{GB} \rightarrow$ Capacidade da porta para o substrato.

A figura 2.8 ajuda a uma fácil visualização da capacidade da porta para o canal e da região de operação.

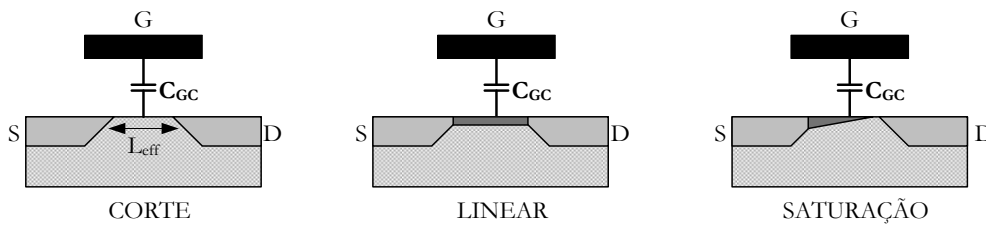


Figura 2.8 – Capacidades da porta para o canal e região de funcionamento do MOSFET

Quando o transistor se encontra na região de corte, como não existe canal, a capacidade que aparece é C_{GB} . Na zona linear, já existe canal de inversão formado entre a fonte e o dreno, logo $C_{GB} = 0$ e a capacidade C_{GC} divide-se igualmente entre o dreno e a fonte, isto é $C_{GD} = C_{GS} = C_{GC}/2$. Por fim na região de saturação o canal está “estrangulado” deixando assim de existir simetria, logo a capacidade entre a porta e o dreno e a porta e o substrato é aproximadamente zero, ficando assim toda a capacidade C_{GC} distribuída entre a porta e a fonte sendo o seu valor cerca de $(2/3)C_{ox}WL$ [18].

Região de Operação	C_{GB}	C_{GS}	C_{GD}	C_{GC}	C_G
Corte	$C_{ox}WL$	0	0	$C_{ox}WL$	$C_{ox}WL+2C_oW$
Linear	0	$C_{ox}WL/2$	$C_{ox}WL/2$	$C_{ox}WL$	$C_{ox}WL+2C_oW$
Saturação	0	$(2/3)C_{ox}WL$	0	$(2/3)C_{ox}WL$	$(2/3)C_{ox}WL+2C_oW$

Tabela 2.1 - Distribuição das capacidades do canal do MOSFET nas diferentes regiões

A tabela 2.1 mostra um sumário do valor das capacidades em cada região de funcionamento.

De referir que em alguma literatura o valor $(2/3)C_{ox}WL$ é por vezes justificado, dizendo que o comprimento do canal na região de *pinch off* é $1/3 L$ e que o comprimento efectivo do canal é $2/3 L$. Esta explicação não é correcta uma vez que o factor $2/3$ resulta de uma derivação matemática[3].

2.3.2 Capacidades Estruturais (Capacidades de Overlap)

As capacidades estruturais aparecem unicamente devido à estrutura do transístor e por isso são as únicas que são lineares. A porta de um transístor MOS é completamente isolada do canal de condução através de um óxido isolante (SiO_2 tipicamente), cuja capacidade por unidade de área é igual a $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$. É usual ter um C_{ox} elevado, para um melhor comportamento I-V.

Idealmente a difusão da fonte e do dreno devem acabar quando começa o óxido da porta mas, como é ilustrado na figura 2.9, na prática devido aos processos de fabrico, isso não acontece. De facto nos transístores reais a difusão da fonte e do dreno está por baixo do óxido da porta num factor x_d denominado difusão lateral.

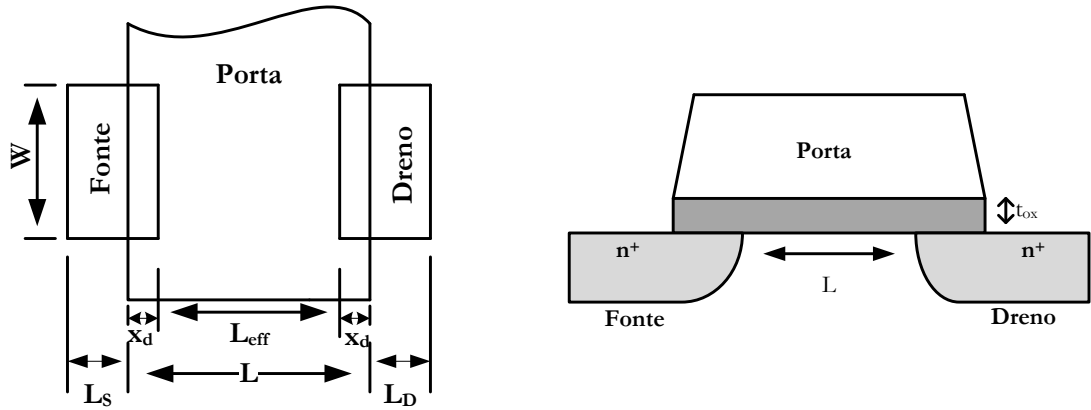


Figura 2.9 - Capacidade de overlap do MOSFET, vista de cima (esquerda) e vista de corte (direita)

Como consequência directa deste facto temos uma variação do comprimento efectivo do canal em $\Delta L = 2 x_d$ [18] passando assim de L para $L_{eff} = L - \Delta L$, e consequente aparecimento de capacidades parasitas entre porta e a fonte (C_{GSO}) e entre a porta e o dreno (C_{GDO}), denominadas capacidades de *overlap*. Como já foi referido estas capacidades são lineares sendo o seu valor:

$$C_{GSO} = C_{GDO} = C_{ox}x_dW = C_OW \quad (2.12)$$

Tendo em vista uma redução nas capacidades de *overlap*, tem-se estudado a possibilidade de construir os MOSFET's na vertical usando um processo denominado FILOX (*fillet local oxidation*) [11] que tem a vantagem de ter uma capacidade de *overlap* porta/dreno menor e ocupar uma menor área de silício quando comparados com os MOSFET's laterais [11, 20, 21].

2.3.3 Capacidade das Difusões

Um outro contributo para as capacidades internas do transistor é as capacidades das junções de dreno - substrato (C_{DB}) e fonte - substrato (C_{SB}). Cada uma delas é causada pela junção PN da base e pela junção lateral. A razão desta distinção deve-se ao facto de durante processo de fabrico a dopagem do substrato perto do canal ser maior

A figura 2.10 mostra uma vista detalhada da junção PN formada entre a fonte e o substrato ou dreno.

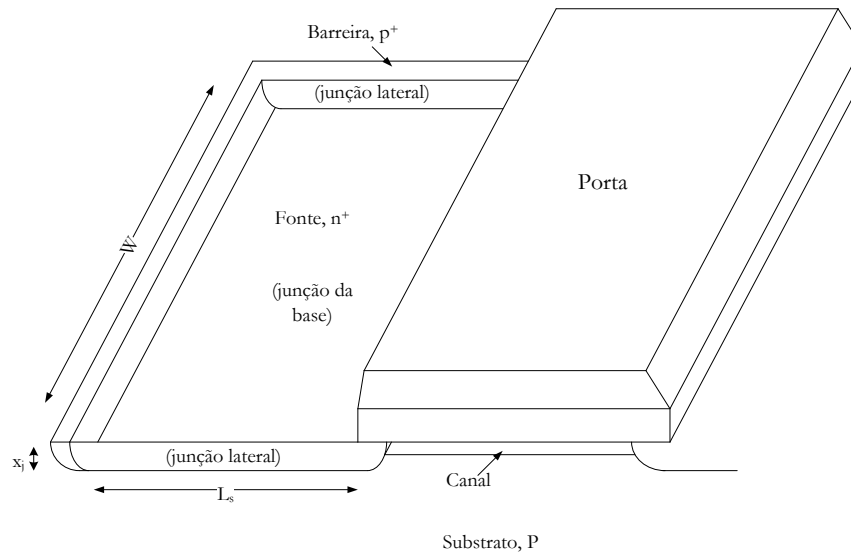


Figura 2.10 – Vista em detalhe da junção da fonte

A capacidade da junção, também conhecida como capacidade de difusão, pode ser dividida em duas capacidades. A primeira é a capacidade da junção da base que é dada pela expressão:

$$C_{bottom} = C_j L_s W \quad (2.13)$$

e a segunda é a capacidade da junção lateral que pode ser calculada a partir de :

$$C_{sw} = C_{jsw}(2L_s + W) \quad (2.14)$$

A capacidade de difusão total é a soma de (2.13) com (2.14) obtendo assim:

$$C_{diff} = C_{bottom} + C_{sw} = C_j L_s W + C_{jsw}(2L_s + W) \quad (2.15)$$

onde C_j é a capacidade da junção por unidade de área e é calculado a partir de :

$C_j = \frac{C_{j0}}{(1 - V_B/\Phi_0)^m}$, C_{j0} é a capacidade da junção quando a polarização é nula, V_B é a tensão aplicada á junção, Φ_0 é o potencial da junção e m é o coeficiente que classifica o tipo de junção quanto à sua dependência com a tensão aplicada, que para a junção da base é cerca de 0.5 (junção abrupta) e para a junção lateral varia entre 0.33 e 0.5 (junção linear) [18].

Faz-se a distinção entre C_j e C_{jsw} , uma vez que as dopagens são diferentes, o que faz variar o C_{j0} e como consequência a capacidade das duas junções são também diferentes.

2.3.4 Modelo de Capacidades do Dispositivo

Todas as capacidades descritas anteriormente podem ser combinadas e escritas segundo um único modelo de capacidades, como pode ser visto na figura 2.11. O modelo das capacidades, foi desenvolvido por Meyer [8] e descreve o comportamento dinâmico dos MOSFET's incluindo apenas os efeitos de primeira ordem. Foi usado também em vários simuladores como o SPICE [9].

O conhecimento deste modelo é muito importante para os desenhadore de circuito.

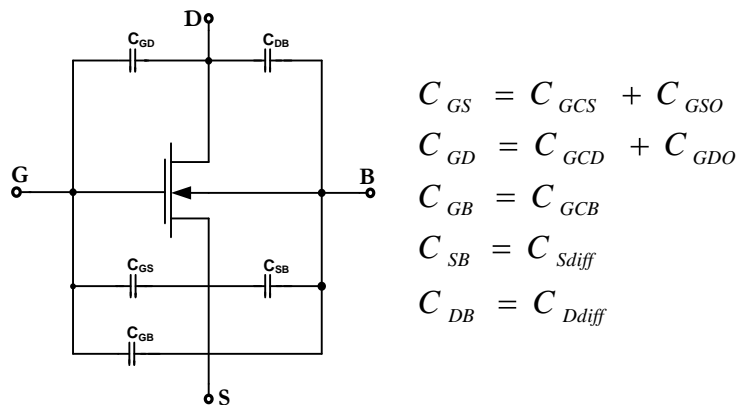


Figura 2.11 – Modelo das capacidades parasitas do MOSFET e respectivas equações

Normalmente o que se procura realmente é a capacidade total da porta do transistor que é calculada a partir da equação 2.16:

$$C_G = (C_{ox}L + 2C_{ox}x_d).W \quad (2.16)$$

A expressão 2.16 representa uma aproximação da capacidade da porta e é útil essencialmente para determinar uma aproximação da capacidade total da porta.

Contudo, apesar de ser simples e suficiente para algumas aplicações, este modelo falha em aplicações cujas propriedades de carga do MOSFET são importantes, tais como circuitos *Silicon-on-Sapphire* (SOS) [10], simulação de SRAM etc.

2.3.5 Comportamento das Ligações

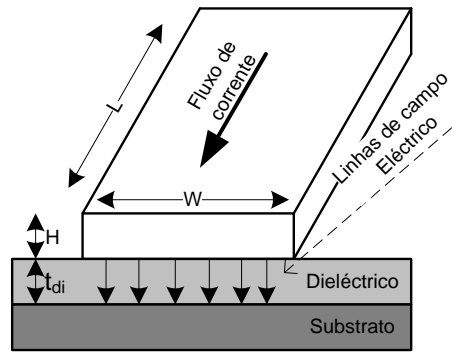


Figura 2.12 – Estrutura de uma ligação

Com o aumento dos circuitos VLSI, as ligações de metal tornaram-se um factor significativo no desempenho de circuitos.

Cada linha de interligação (*wire*) é uma estrutura tridimensional, figura 2.12, em metal e/ou polisilício com variações significativas na sua forma, espessura ou distância para o substrato. Além disso cada ligação pode ser rodeada por outras ligações no mesmo nível ou em níveis diferentes [18]. A figura 2.13 mostra

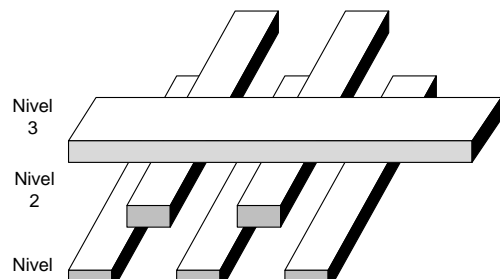


Figura 2.13 - Exemplo de seis linhas de interligação em três níveis diferentes

uma possível situação onde as interligações de três níveis diferentes estão próximas.

Usando em uma primeira análise a geometria da figura 2.8, e assumindo que as ligações formam condensadores de placas paralelas entre si (C_{pp}), se a largura da ligação é

muito maior que a espessura do material isolante, então é possível assumir que as linhas de campo eléctrico são ortogonais às placas do condensador e o seu valor é dado por:

$$C_{pp} = \frac{\epsilon_{di}}{t_{di}} WL \quad (2.17)$$

Onde W e L representam a largura e o comprimento da ligação respectivamente e o t_{di} e ϵ_{di} a espessura do dieléctrico e a sua permitividade. [18].

Hoje este modelo é bastante simples e desadequado da realidade actual, uma vez que a constante redução da espessura do óxido e da distância entre os metais faz com que o campo eléctrico deixasse de ser ortogonal, figura 2.14, e a capacidade entre as paredes laterais das ligações e o substrato deixou de poder ser ignorada.

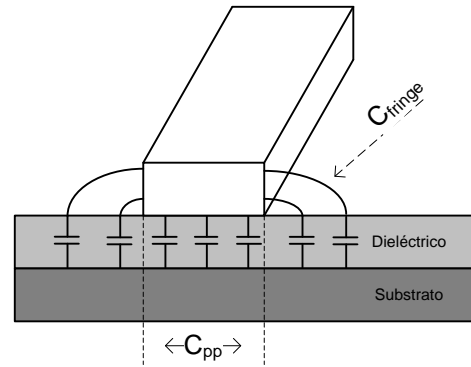


Figura 2.14 - Influência do C_{fringe} nas ligações

Uma forma mais a aproximada para calcular o valor real das capacidades das ligações é dada pela fórmula [5, 17]:

$$C_{wire} = L\epsilon_{di} \left[\frac{W}{t_{di}} - \frac{H}{2t_{di}} + \frac{2\pi}{\ln\left(1 + \frac{2t_{di}}{H} \left[1 + \sqrt{1 + \frac{H}{t_{di}}}\right]\right)} \right] \quad (2.18)$$

2.4 Scaling

Desde a década de 60, altura em que os MOSFET's começaram a ser usados, tem-se assistido a um aumento da integração dos circuitos integrados. Este aumento de integração foi uma consequência directa da evolução dos processos litográficos (fotoquímico), que permitiu a diminuição do comprimento de canal dos dispositivos, possibilitando assim um maior número de transístores por *chip*. Esta diminuição constante das dimensões físicas ficou conhecida como “*scaling*”.

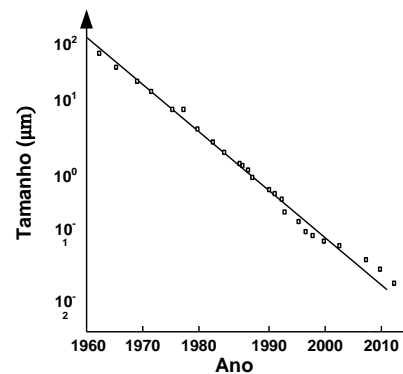


Figura 2.15 - Evolução da diminuição do comprimento do canal

Como se pode observar na figura 2.15, a diminuição mínima dos transístores (*minimum feature size*), tem vindo a decrescer a uma taxa de cerca de 13% por ano [18].

A tecnologia CMOS era uma tecnologia quase insignificante nos anos 60 mas com o aumento da escala de integração impulsionada pelo *scaling* bem como uma reduzida dissipação de potência, tornaram-na a tecnologia dominante até à década de 90.

2.4.1 *Scaling Completo (ou Scaling de Campo Eléctrico constante)*

Neste método de *scaling*, tensões (V_{DD} , V_{th}) e dimensões (W , L , t_{ox}) são reduzidas num mesmo factor, factor S (com $S > 1$). O principal objectivo desta abordagem é uma diminuição dos dispositivos, mantendo um campo eléctrico idêntico entre o dispositivo original e o reduzido, evitando assim o aparecimento de novos efeitos de segunda ordem [15].

Este modelo tem a vantagem de oferecer uma maior densidade (menos área ocupada), maior performance, isto é tempos de propagação menores (redução em um factor S), e uma redução na potência consumida num factor de S^2 [22].

Como desvantagem, este método tem o facto de que a redução nas tensões de limiar provoca um aumento da condução na região de sublimiar e consequentemente um aumento das correntes de fuga.

De facto o *scaling* completo não é um método muito útil na prática, uma vez que as tensões de alimentação são reduzidas no mesmo factor da redução das dimensões do transístor, impossibilitando assim estabelecer um valor de tensão padrão para a interface com o *chip*.

2.4.2 *Scaling por Tensões Fixas*

O *scaling* por tensões fixas (quinta coluna da tabela 2) surgiu uma vez que os diferentes componentes do chip interagem entre si, portanto interessa ter um valor padrão nas tensões de alimentação, poupando assim no custo em fontes de tensão.

Como resultado as tensões não são decrementadas com a diminuição das dimensões dos MOSFET, que são reduzidas por um factor S [19]. Convém referir que a espessura do óxido t_{ox} não deverá ser reduzida no mesmo factor com o objectivo de não

aumentar o campo eléctrico, provocando degradação na mobilidade dos portadores de carga.

Criaram-se assim valores padrão para as tensões das fontes de alimentação e para os níveis lógicos. Na década de 90 este foi o modelo seguido, e a tensão utilizada era 5V. Actualmente a tensão utilizada cifra-se entre 3.3V e 1V com a introdução das tecnologias CMOS de 0.5 μ m até 90nm.

Este modelo tem a desvantagem de consumir bastante mais potência, já que a sua densidade de corrente aumenta de uma forma quadrática, S^2 , o que vai originar um “*breakdown*” no óxido, também conhecido como *second breakdown*, trata-se de uma falha irreversível nos semicondutores, e é caracterizado por um aumento de corrente no óxido que vai originar *hotspots*, que o destroem. Estes efeitos tornaram insustentável o modelo de tensões fixas.

2.4.3 Scaling Geral

Uma vez que algumas barreiras eram difíceis de ultrapassar tais como: tensões intrínsecas do MOSFET que são parâmetros do material e não podem ser reduzidas (exemplo o potencial do Fermi φ_F); e a diminuição da tensão de limiar que tornava quase impossível desligar totalmente o dispositivo, obrigaram a usar um modelo de scaling mais geral (quarta coluna da tabela 2) em que as dimensões e as tensões são reduzidas com parâmetros diferentes, S e U respectivamente. Quando $U = 1$ o método de *scaling* geral fica reduzido em um *scaling* por tensões fixas [18].

Este método é o melhor dos três uma vez que aproveita o melhor de cada um dos anteriores, isto é, reduz a potência dissipada e melhora a nível de densidade de potência.

De referir também que sob o ponto de vista económico o *scaling* geral é mais vantajoso pois assegura uma compatibilidade ao nível das tensões com tecnologias anteriores.

A tabela 2 demonstra não só como os parâmetros do transistor variam consoante o método de *scaling* utilizado, mas também a relação existente entre alguns desses parâmetros.

Parâmetro	Relação	Scaling Completo	Scaling Geral	Scaling por Tensões Fixas
W, L, t_{ox}		$1/S$	$1/S$	$1/S$
V_{DD}, V_T		$1/S$	$1/U$	1
Área	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	$1/S$	$1/S$	$1/S$
N_A, N_D	V/W_{dept}^2	S	S^2/U	S^2
K_n, k_p	$C_{ox}W/L$	S	S	S
Densidade de Corrente	$I_{sat}/Área$	S	S^2/U	S^2
R_{on}	V/I_{sat}	1	1	1
t_p	$R_{on}C_{gate}$	$1/S$	$1/S$	$1/S$
P_{av}	I_{sat}/V	$1/S^2$	$1/U^2$	1
Densidade de Potência	$P/Área$	1	S^2/U^2	S^2

Tabela 2.2– Influência do scaling nos diferentes parâmetros do MOSFET

2.5 Impacto do Scaling nos Dispositivos e Resolução de Alguns Problemas

O progressivo *scaling* da tecnologia fez variar não só as dimensões físicas do transistor, mas também o seu comportamento, tornando-se por vezes complicado depreender qual a zona de funcionamento em que encontra o transistor. A principal conclusão a retirar é que os transístores de canal curto têm uma zona de saturação maior mas ao mesmo tempo têm menor capacidade de distribuir corrente já que antes de saturarem por *pinch-off*, saturam por velocidade de deriva [3, 18]. O aumento da importância das capacidades parasitas tornou-se também um problema que forçou os desenhadores de circuitos lógicos a desenvolver novas estruturas para os transístores.

Como já referido anteriormente, o *Fillox*, foi uma das novas estruturas adoptadas para combater a manifestação destes efeitos [11], mas não foi a única. Um outro exemplo foio aparecimento das estruturas SOI (*Silicon-on-Insulator*) que se destaca por ter um BOX

(*Buried Oxide*), que é uma deposição de uma fina camada de óxido no substrato do dispositivo e serve para isolar dielectricamente o transistor [17]. Este isolamento permite um melhor controlo sobre as correntes de fuga. De entre as várias estruturas SOI implementadas, destacam-se o transistor GAA (Gate All-Around), o FinFET, o DGSOI (Double-Gate SOI) e o DTMOS (Dynamic Threshold MOS) [21].

Vários investigadores crêem que o nanotubo de carbono irá ser o substituto do MOSFET tradicional, já que se trata de um material mais forte que o aço, mais leve que o alumínio e melhor condutor que o cobre. Acredita-se que este nanotubo irá desempenhar o papel de canal do transistor e estará isolado do restante dispositivo, o que irá aumentar bastante o desempenho do transistor de silício [26].

2.6 Sumário

O componente base para a construção de qualquer circuito digital é o transistor do tipo MOS. Neste capítulo foram discutidos os vários efeitos secundários presentes no MOSFET, dando ênfase aos efeitos denominados de canal curto como a saturação de velocidade de deriva. De seguida falou-se sobre as capacidades intrínsecas do MOSFET, já que estas capacidades influenciam o comportamento dinâmico do dispositivo.

Foram apresentados também vários modelos de *scaling* e os efeitos que causam na física do dispositivo. Por fim é explicada a necessidade de adoptar novas estruturas e técnicas de forma a combater os vários efeitos que têm vindo a surgir nos dispositivos de canal curto.

Capítulo 3 – Desenho de Portas Lógicas CMOS

A velocidade e o equilíbrio dos circuitos CMOS sempre foi uma prioridade para os desenhadores de circuitos lógicos. No entanto, grande parte das técnicas de desenho foi desenvolvida quando os transístores tinham um comportamento bem definido.

Uma escolha menos correcta do rácio entre as dimensões do PMOS e do NMOS, o efeito de *Miller* provocado pela carga e pelo *driver* bem como as capacidades parasitas que se tornam mais relevantes em tecnologias de canal curto, são alguns efeitos que desequilibram os tempos de propagação.

Assim a métrica usada uma boa métrica é o desequilíbrio, já que relaciona o t_{pHL} e t_{pLH} .

Neste capítulo vai começar por se descrever como se constroem as portas CMOS, explicando em pormenor a porta NAND e a porta NOR, de seguida vai abordar-se o efeito da capacidade de *Miller* tanto na porta *driver* como na porta de carga.

3.1 Portas CMOS – Conceito

Nas portas CMOS, a função lógica é formada por dois circuitos complementares conhecidos por PUN (*Pull-Up Network*) e PDN (*Pull-Down Network*), como mostrado na figura 3.1. Estas duas redes são construídas com o intuito de apenas uma delas conduzir isto é para cada combinação de entrada, apenas uma é activada.

O PUN é construído exclusivamente com transístores PMOS e a sua função é estabelecer uma ligação entre a saída e V_{DD} , para qualquer combinação de entrada em que $F = '1'$ [18]. Similarmente o PDN é formado unicamente por transístores NMOS e o seu objectivo é ligar a saída à massa para todas as combinações em que $F = '0'$.

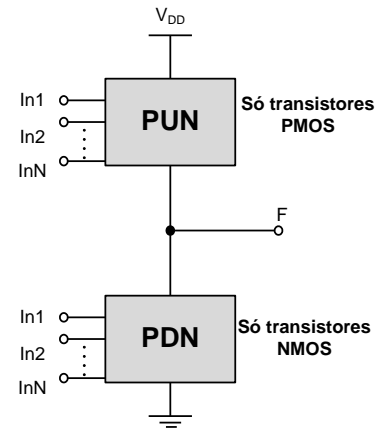


Figura 3.1– Estrutura de uma porta lógica CMOS

A regra básica na construção das redes de transístores no PDN é: NMOS ligados em série correspondem à função NAND, isto é com todas as entradas activas a série de transístores conduz, logo o valor á saída é '0'. Por outro lado a uma sequência de transístores em paralelo no PDN criam a função NOR, isto porque basta uma entrada estar activa para na saída se obter o valor lógico '0' [18].

A construção da rede de PMOS é dual, isto é, transístores em paralelo formam a função lógica NAND e em série a NOR.

Uma vantagem muito importante das portas CMOS é o facto de todas as propriedades conhecidas para o inversor simples serem válidas ($V_{OH} = V_{DD}$, $V_{OL} = 0V$, potência estática nula etc.) para as restantes portas.

3.2 Porta NAND

A figura 3.2, mostra uma porta NAND de duas entradas. O PDN desta porta é constituído por dois dispositivos NMOS em série e como já foi referido anteriormente, o PUN é o dual, isto é, são dois PMOS em paralelo.

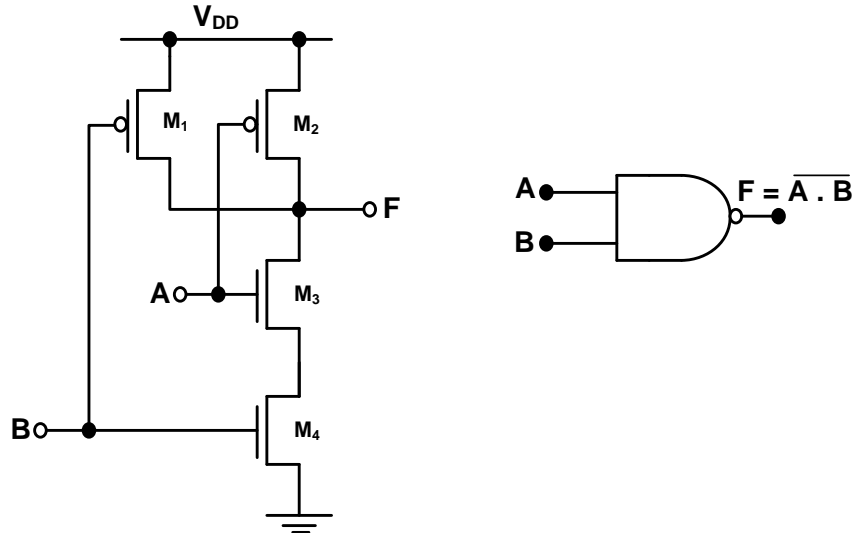


Figura 3.2 - Esquema eléctrico, símbolo e função lógica de uma NAND de duas entradas

Considerando o ramo série se ambas as *gates* estiverem no estado lógico *high*, significa que os transístores estão activos e conduzem corrente, criando assim um caminho de condução para a massa e portanto o nível lógico na saída será '0' ($V_{OL} = 0V$). Porém se um ou os dois NMOS estiverem desactivados, com as *gates* no estado *low*, deixa de existir caminho para a massa e pelo menos um PMOS estará em condução, tendo assim a tensão de saída igual a V_{DD} ($V_{OH} = V_{DD} = '1'$).

3.2.1 Característica Estática da NAND2

A função característica DC (VTC – *Voltage Transfer Characteristic*) depende da combinação de entradas. Há três possibilidades de trocar o estado da porta lógica de *high* para *low*:

- (a) $V_A = V_B = 0 \rightarrow 1$
- (b) $V_A = V_{DD}$ e $V_B = 0 \rightarrow 1$
- (c) $V_B = V_{DD}$ e $V_A = 0 \rightarrow 1$

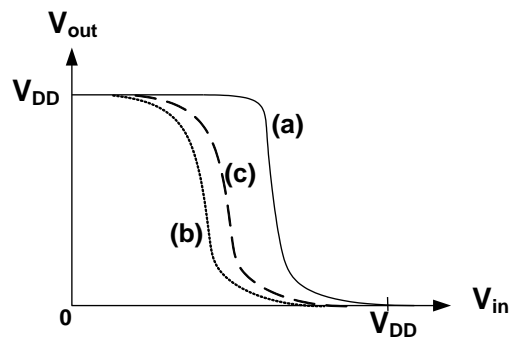


Figura 3.3 – Várias VTC's de uma NAND2

As diferentes combinações de entradas mostram diferenças significativas. A grande variação entre o caso (a) e os restantes (b e c) pode ser explicado pelo facto de no início da

transição ambos os transístores do PUN se encontrarem activos ($A=B=0$) representando um forte *pull-up*. Nos restantes casos apenas um PMOS se encontra activo, portanto temos um *pull-up* mais fraco, e por conseguinte as curvas estão mais encostadas à esquerda. A diferença entre (b) e (c), embora pequena resulta da variação de V_{th} de M_3 devido ao efeito de corpo neste transístor.

$$V_{thA} = V_{th0} + \gamma(\sqrt{|2\Phi_F| + V_{DSM4}} - \sqrt{|2\Phi_F|}) \quad (3.1)$$

$$V_{thB} = V_{th0} \quad (3.2)$$

3.2.2 Tempos de Propagação da NAND2

Para o cálculo dos tempos de propagação é fundamental saber o valor da capacidade de saída da porta lógica (C_{out}), calculado usando 3.3:

$$C_{out} = C_{FET} + C_L \quad (3.3)$$

Onde,

$$C_{FET} = C_{GDM3} + C_{GDM1} + C_{GDM2} + C_{DBM3} + C_{DBM1} + C_{DBM2} \quad (3.4)$$

Em que C_{GDM3} , C_{GDM1} e C_{GDM2} são as capacidades da porta para o dreno de M_3 , M_1 e M_2 respectivamente e C_{DBM3} , C_{DBM1} e C_{DBM2} são por esta ordem as capacidades do dreno para o corpo do M_3 , M_1 e M_2 e:

$$C_L = C_{line} + C_{FO} \quad (3.5)$$

Onde C_{line} é a capacidade introduzida pelas ligações e C_{FO} é a capacidade que aparece devido às portas de carga.

De referir que C_{FET} é dependente da(s) entrada(s) onde é aplicado o sinal. Por exemplo se $V_B = V_{DD}$ e $V_A = 0 \rightarrow 1$, então C_{GDM3} e C_{GDM2} aparecem multiplicada por um factor na equação 3.4.

Considerando a transição *low-to-high*, isto é, inicialmente $V_{out} = 0$ V, que implica que tanto V_A como V_B estão inicialmente ligados a V_{DD} . Se A, B ou as duas entradas trocarem para o estado lógico '0', C_{out} começa a ser carregado devido ao PUN que fica activo.

Usando a expressão 3.6 [18, 23]:

$$t_{pLH} = \ln(2)\tau_p \quad (3.6)$$

com:

$$\tau_p = R_p C_{out} \quad (3.7)$$

caso apenas um PMOS esteja ligado ou

$$\tau_p = R_p/2 C_{out}, \quad (3.8)$$

se os dois PMOS estiverem activos

τ_p é a constante de tempo do circuito e R_p é a resistência equivalente de um transistor PMOS. Como é possível depreender desta análise de primeira ordem, o pior caso acontece quando apenas um transistor do PUN se encontra em condução.

Existe apenas uma combinação que obriga a porta NAND a transitar de *high* para *low* que é quando as entradas dos dois transistores do ramo série se encontram ligados a VDD, a equação 3.6, continua a poder aplicar-se mas τ_p , calcula-se agora a partir de:

$$\tau_p = 2R_n C_{out} \quad (3.9)$$

3.3 Porta NOR

Uma NOR de duas entradas, em CMOS, é construída usando dois pares complementares como ilustra a figura 3.4.

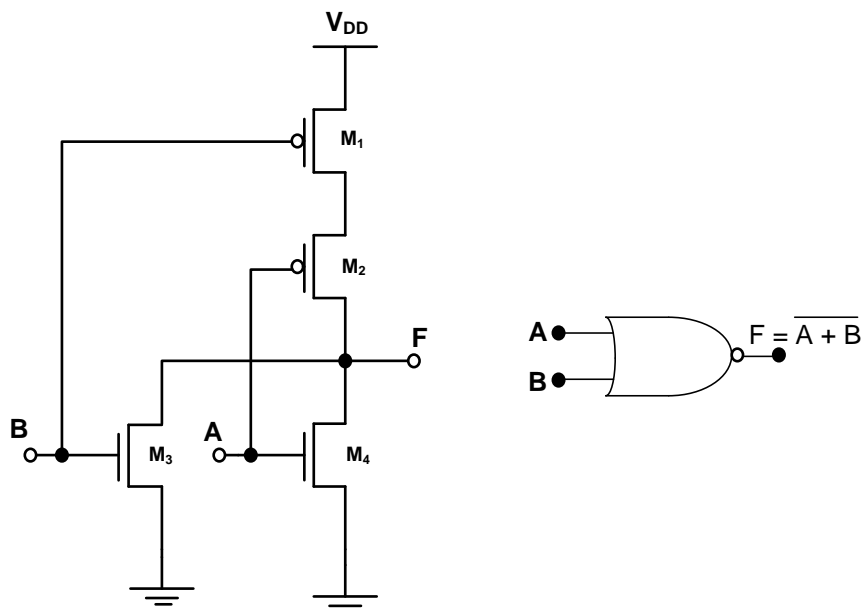


Figura 3.4 - Esquema eléctrico, símbolo e função lógica de uma NOR de duas entradas

Para perceber a operação da porta vai-se examinar os estados de condução dos transístores para as diferentes tensões de entrada. Se $V_A = V_{DD}$ ou $V_B = V_{DD}$, então o NMOS-A ou NMOS-B estão activos e o PMOS-A ou PMOS-B cortado, logo o valor lógico à saída será '0'. Se $V_A = V_B = V_{DD}$, os NMOS estão activos e os PMOS cortados, portanto a saída também irá ser zero. Portanto a única combinação de entradas em que $F=1$, ocorre quando $V_A = V_B = 0V$, ambos os PMOS estão activos, o que gera um caminho até V_{DD} .

3.3.1 Característica estática da NOR2

A tensão DC da saída da porta NOR depende exclusivamente das entradas e da forma como elas variam. Tal como na NAND2, temos três diferentes VTC possíveis como mostra a figura 3.5:

- (a) $V_B = 0 \quad V_A = 0 \rightarrow 1$
- (b) $V_A = V_B = 0 \rightarrow 1$
- (c) $V_A = 0$ e $V_B = 0 \rightarrow 1$

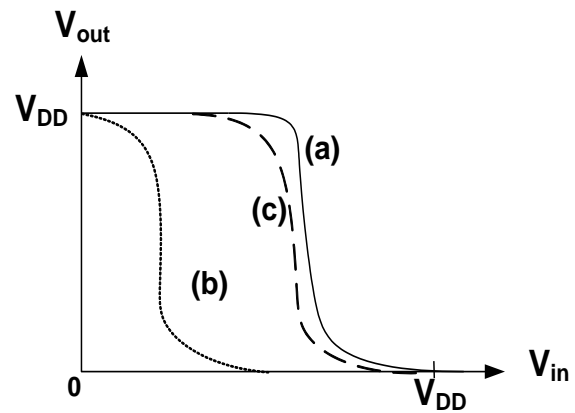


Figura 3.5 -Diferentes VTC's de uma NOR2

É possível observar que cada caso é caracterizado por VTCs diferentes, mas contrariamente à NAND2, quando só varia uma das entradas, a curva fica mais à direita no gráfico.

3.3.2 Tempos de Propagação da NOR2

Para estimar os tempos de propagação da porta é necessário saber o valor de C_{out} . Este pode ser calculado a partir da equação 3.3, C_{FET} é agora calculado usando a expressão 3.10:

$$C_{FET} = C_{GD_{M3}} + C_{GD_{M4}} + C_{GD_{M2}} + C_{DB_{M3}} + C_{DB_{M4}} + C_{DB_{M2}} \quad (3.10)$$

Com C_{GD} e C_{DB} como capacidades da porta para o dreno e capacidade do dreno para o corpo respectivamente dos vários transístores.

O pior caso para o tempo de descarga do C_{out} , ocorre quando apenas um NMOS conduz. Esta situação é equivalente à descarga de um simples inversor.

Para calcular o t_{pLH} , assume-se que as entradas A e B se encontram no estado lógico '0', portanto apenas os PMOS conduzem sendo a constante de tempo do circuito RC:

$$\tau_p = (R_{pA} + R_{pB})C_{out} \quad (3.11)$$

Onde R_{pA} e R_{pB} é a resistência equivalente dos PMOS. A expressão para calcular o tp_{LH} é :

$$tp_{LH} = \ln(2)\tau_p \simeq 0.69\tau_p \quad (3.12)$$

Expressão aproximada e semelhante à calculada para a NAND2.

Como referido anteriormente, o dimensionamento dos transístores das várias portas lógicas, tem como base um inversor de referência

3.4 Comparação entre as portas NAND e NOR

Quando se dimensiona os transístores das portas lógicas parte-se de um inversor de referência, isto é, o circuito é desenhado para que os piores tempos de propagação da porta sejam iguais aos tempos de propagação do inversor de referência. Assim os NMOS da porta NAND-N (N entradas) têm uma relação (W/L), N vezes superior à relação (W/L) do inversor. Por outro lado nos PMOS as relações (W/L) mantêm-se iguais. De uma forma dual os NMOS de uma porta NOR-N mantêm a mesma relação (W/L) do inversor de referência, enquanto os PMOS têm uma relação (W/L) N vezes superior.

A NAND e o NOR são portas simples de implementar na tecnologia CMOS. Contudo para o mesmo número de entradas e tamanhos dos dispositivos, a porta NAND tem melhor resposta dinâmica que a porta NOR. Este facto pode ser explicado através dos tempos de propagação.

A série de transístores é um factor limitativo. Na NAND o tp_{HL} é determinado através da resistência formada pela cadeia de NMOS, por outro lado tp_{LH} na NOR é calculado a partir da resistência formada pela cadeia de PMOS.

Como é possível concluir da expressão 3.13,

$$R = \frac{1}{K' \left(\frac{W}{L}\right) (V_{DD} - V_{th})} \quad (3.13)$$

a resistência de um transistor depende das suas dimensões e como $(K'_n/K'_p) > 1$, para as mesmas dimensões de PMOS e NMOS, então $R_p > R_n$. Partindo desta premissa é lógico afirmar que a cadeia de NMOS é mais rápida que a cadeia de PMOS. Deste modo, para a mesma área de silício as portas NAND são preferíveis quando comparadas com a NOR [18, 23].

3.5 Efeito de Miller

A figura 3.6 mostra um inversor CMOS com a respectiva capacidade de Miller (CM).

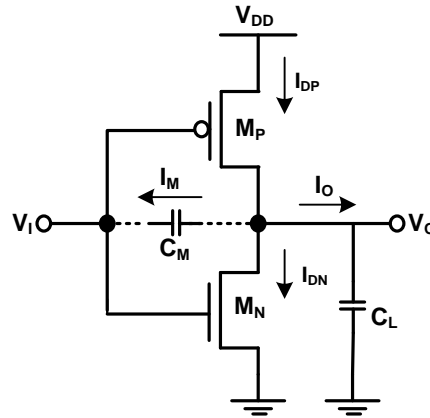


Figura 3.6 - Esquema básico de um inversor CMOS, com a respectiva capacidade de Miller

3.5.1 Modelo de carga simples

Fazendo uma primeira análise deste circuito, recorrendo às leis de *Kirchoff* e ignorando a capacidade de Miller, tem-se:

$$I_{DP} = I_{DN} - I_O \quad (3.14)$$

Utilizando a expressão da corrente num condensador, esta equação torna-ser:

$$I_{DP} - I_{DN} = C_L \frac{\partial V_o}{\partial t} \quad (3.15)$$

C_L carrega, através do PMOS (tp_{IH}), e descarrega, através do NMOS (tp_{HL}). Assim quando a entrada varia de V_{DD} até à massa, o M_N está cortado. C_L carrega de 0V até V_{DD} através de M_p que está saturado logo a corrente é:

$$I_{DP} = K_P (V_{DD} - |V_{TP}|)^2 \quad (3.16)$$

V_o aumenta até atingir o valor de $|V_{TP}|$. Neste ponto M_p muda o modo de operação para o estado linear demorando um tempo t_1 até isso acontecer. Combinando a

equação 3.16 (com $I_{DN}=0$ uma vez que o NMOS está cortado) com a equação 3.17, obtém-se uma expressão para o valor de t_1 :

$$t_1 = \frac{|V_{tp}|C_L}{K_p(V_{DD}-|V_{tp}|)^2} \quad (3.17)$$

A partir de t_1 , o C_L começa a ser carregado com uma corrente de:

$$i_{DP} = K_p[2(V_{DD} - |V_{TP}|)(V_{DD} - V_O) - (V_{DD} - V_O)^2] \quad (3.18)$$

O tempo que demora a carregar o condensador de $|V_{TP}| < V_O < V_{DD}/2$ é chamado de t_2 e é calculado a partir de :

$$t_2 = \frac{C_L}{2K_p(V_{DD}-|V_{TP}|)} \ln \frac{3V_{DD}-4|V_{TP}|}{V_{DD}} \quad (3.19)$$

Por isso o tempo de propagação de *low* para *high* é $t_{pLH} = t_1 + t_2$ logo:

$$t_{pLH} = \frac{C_L}{2K_p(V_{DD}-|V_{TP}|)} \left[\frac{2|V_{TP}|}{V_{DD}-|V_{TP}|} + \ln \frac{3V_{DD}-4|V_{TP}|}{V_{DD}} \right] \quad (3.20)$$

Similarmente é possível obter o tempo *high-to-low* calculado durante a descarga do condensador:

$$t_{pHL} = \frac{C_L}{2K_N(V_{DD}-V_{TN})} \left[\frac{2V_{TN}}{V_{DD}-V_{TN}} + \ln \frac{3V_{DD}-4V_{TN}}{V_{DD}} \right] \quad (3.21)$$

Para se obter um inversor simétrico, isto é, com $t_{pHL}=t_{pLH}$, é necessário que $K_p=K_N$, e $|V_{TP}|=V_{TN}$ [23].

3.5.2 Modelo Incluindo a Capacidade de Miller do Inversor

A capacidade parasita de *Miller*, que aparece representada na figura 3.6 é formada pelas capacidades C_{GD} do NMOS e do PMOS.

$$C_{GD} = C_{GDoverlap} + C_{GDcanal} \quad (3.22)$$

Dependendo da transição lógica na saída C_M é carregado ou descarregado. Numa perspectiva de tempos de propagação, a capacidade de Miller introduz significativas não-linearidades que podem causar a mudança da região de operação dos dois transístores.

Assim tendo em conta C_M , obtém-se o seguinte modelo:

$$I_{DP} = I_{DN} + I_M + I_O \quad (3.23)$$

$$\Leftrightarrow I_{DP} - I_{DN} = (C_L + C_M) \frac{\partial V_O}{\partial t} - C_M \frac{\partial V_I}{\partial t} \quad (3.24)$$

Derivando a equação 3.24 e assumindo que o transístor NMOS se encontra sempre a conduzir na zona de saturação ou seja I_{DN} é uma constante, e $I_{DP} = 0$, e assumindo que a tensão de entrada tem transições instantâneas chega-se á expressão da tensão de saída:

$$V_{out}(t) = \frac{C_L + 2C_M}{C_L + C_M} V_{DD} - \frac{I_{DP}}{C_L + C_M} t \quad (3.25)$$

O t_{pHL} é calculado quando $V_{out}(t)$ atinge $V_{DD}/2$, assim:

$$t_{pHL} = \frac{C_L + 3C_{Mdescida}}{2} \cdot \frac{V_{DD}}{I_{DN}} \quad (3.26)$$

Analogamente o tempo de *low* para *high* é calculado usando a expressão 3.25

$$t_{pHL} = \frac{C_L + 3C_{Msubida}}{2} \cdot \frac{V_{DD}}{I_{DP}} \quad (3.27)$$

As expressões 3.26 e 3.27 revelam a importância que a capacidade de *Miller* tem na análise dos tempos de propagação em inversores CMOS [24].

Existem significativas variações da capacidade de *Miller* durante as transições lógicas do inversor. Durante cada transição o NMOS e o PMOS operam em regiões diferentes (corte, linear e saturado). Na transição de *high* para *low*, o PMOS opera inicialmente na região linear, isto é, máximo C_{GD} , enquanto o NMOS opera na região de saturação (C_{GD} mínimo). Durante a transição, o PMOS começa a cortar gradualmente e o NMOS entra a região linear. A grande diferença na capacidade total de *Miller* deve-se ao facto dos transístores PMOS serem maiores comparativamente aos NMOS [24].

3.5.3 Modelo da Capacidade de Miller do Inversor Driver e da Carga

A figura 3.7 mostra um circuito com um inversor driver e um inversor de carga. Para uma completa análise do circuito é inevitável estudar a evolução temporal de V_a (saída do inversor de carga) em ordem ao valor de V_o (saída do inversor de driver).

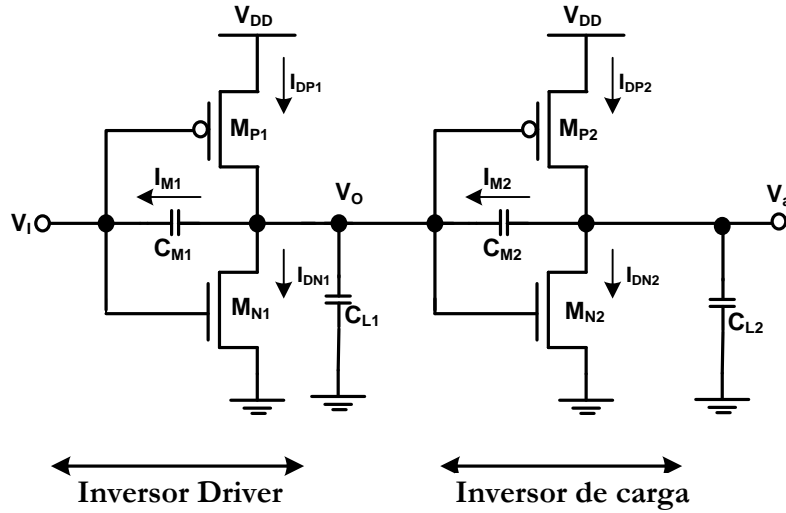


Figura 3.7 – Inversor CMOS tendo outro inversor CMOS como carga

Assim o circuito pode ser descrito recorrendo a duas equações diferenciais:

$$\begin{cases} (C_{M1} + C_{L1} + C_{M2}) \frac{dV_o}{dt} - C_{M2} \frac{dV_a}{dt} = I_{DP1} - I_{DN1} + C_{M1} \frac{dV_i}{dt} \\ -C_{M2} \frac{dV_o}{dt} + (C_{M2} + C_{L2}) \frac{dV_a}{dt} = I_{DP2} - I_{DN2} \end{cases} \quad (3.28)$$

onde, I_{DPK} e I_{DNK} representam as correntes de dreno dos transístores M_{PK} e M_{NK} [25].

Uma análise cuidadosa à equação 3.28 mostra uma dependência não linear no V_o entre I_{DP2} e I_{DN2} , o que torna a segunda equação diferencial de 3.28 não linear. O mesmo raciocínio pode ser feito para I_{DP1} e I_{DN1} e V_i , mas neste caso a equação associada é linear em V_o e V_a . Contudo o sistema é não linear o que torna bastante complicado encontrar uma solução analítica para V_o sem assumir simplificações.

Para um inversor simples é possível que durante a transição *high* para *low* (*low* para *high*) apenas o NMOS (PMOS) conduz e é capaz de descarregar (carregar) a capacidade de carga [24, 27]. Partindo desta afirmação e considerando que o inversor *driver* transita do estado ‘1’ para o estado ‘0’, então o inversor de carga muda o seu estado de ‘0’ para ‘1’ e a equação 3.28 pode ser simplificada em:

$$\begin{cases} (C_{M1} + C_{L1} + C_{M2}) \frac{dV_o}{dt} - C_{M2} \frac{dV_a}{dt} = -I_{DN1} + C_{M1} \frac{dV_i}{dt} \\ -C_{M2} \frac{dV_o}{dt} + (C_{M2} + C_{L2}) \frac{dV_a}{dt} = I_{DP2} \end{cases} \quad (3.29)$$

A equação 3.31 mostra que V_o depende não apenas das características do NMOS do *driver*, mas também do PMOS da carga. Assim, o t_{pLH} do driver é também dependente do transístor PMOS da carga [25].

Assumindo que a solução do V_o é uma função continua no tempo, então V_a pode ser obtido na segunda equação de 3.30 em que $V_a(t) = f(V_o(t))$ representa a solução. $f(V_o(t))$ pode ser interpretado como a VTC do inversor de carga. Aplicando esta solução na primeira equação de 3.30 e usando a propriedade da derivada de funções compostas obtém-se:

$$\left[C_{M1} + C_{L1} + C_{M2} \left(1 - \frac{df}{dV_o} \right) \right] \frac{dV_o}{dt} = I_{DP1} - I_{DN1} + C_{M1} \frac{dV_i}{dt} \quad (3.30)$$

Para frequências de operação moderadas, $f(V_o(t))$ pode ser aproximada a uma VTC estática de um inversor, assim df/dV_o pode ser visto como um ganho em tensão. Num inversor, o ganho de tensão é sempre negativo e o seu valor mínimo ocorre durante a transição entre estados. Como consequência, o efeito de C_{M2} pode vir muito amplificado pelo ganho em tensão do inversor de carga. Este é, de facto, a manifestação do efeito de Miller da carga, quando a carga é um inversor (ou uma outra porta) em vez de uma carga linear simples [25].

O valor de V_{th} têm uma dependência quadrática ou linear (para a saturação ou região linear) [18] com I_{DS} tendo assim um impacto no ponto de transição da porta lógica. Correspondentemente uma correcção no caminho resistivo do PUN e PDN afecta linearmente o valor da corrente entre o dreno e a fonte em ambas as regiões de operação do MOSFET. Este ajuste torna este estudo mais preciso, mas aumenta significativamente a complexidade deste mesmo estudo se o declive da curva negligencia a corrente de curto-circuito e as capacidades de acoplamento ou se consideram ambas.

O VM da porta também é um parâmetro de estudo bastante válido, uma vez que é calculado com base nas mobilidades dos transístores devido a isso irão ser feitos testes baseados neste parâmetro.

Parâmetros como as mobilidades dos portadores de carga, dimensões do transístor do inversor de driver e de carga, variações nas tensões de threshold são alguns factores que influenciam $t_{p_{HL}}$ e $t_{p_{LH}}$ na saída do inversor de driver. Assim, a conceito de um inversor com base na configuração simples da figura 3.6, a fim de apresentar tempos de propagação iguais, não implica que o mesmo comportamento seja válido para as condições de carga diferentes como irá ser mostrado no capítulo 4.

3.6 Sumário

O equilíbrio dos circuitos CMOS é uma das grandes preocupações dos desenhadores de circuitos digitais, já que circuitos desequilibrados podem destruir o sinal e assim o bom funcionamento do circuito. Assim neste capítulo discutem-se técnicas de construção das portas lógicas CMOS e fez-se uma comparação entre elas.

É exposto o problema da capacidade de *Miller*, tanto na porta de driver como na porta de carga, e é apresentado um modelo matemático sobre esta capacidade.

Capítulo 4 – Resultados

Devido à grande evolução das tecnologias CMOS bem como os problemas que têm aparecido devido aos efeitos secundários, obrigou os projectistas de circuitos integrados a recorrer a utilização de ferramentas avançadas de CAD (*Computer-Aided Design*). Neste capítulo é explicado, inicialmente, os vários passos para o desenvolvimento das portas lógicas usadas neste estudo. De seguida, são apresentados os vários esquemáticos usados na simulação e por fim expõem-se os resultados e as várias conclusões retirados deles.

4.1 Tecnologia CMOS Utilizada

O trabalho desenvolvido baseou-se numa tecnologia específica, sendo o *design kit* utilizado o AMS (*Austria MicroSystems*) de 350 nm.

As dimensões mínimas de cada MOSFET são $W_{min} = 0.4\mu\text{m}$ e $L_{min} = 0.35\mu\text{m}$, a tensão de alimentação suportada e utilizada na realização das várias experiências foi de 3.3V.

De notar que na elaboração de uma das experiências utilizou-se um inversor ideal fornecido pela própria biblioteca.

4.2 Método de Desenvolvimento

O diagrama da figura 4.1 oferece uma visão simplificada da metodologia e a sequência de etapas para o desenvolvimento das portas lógicas. A criação de cada porta lógica passa por duas fases principais: sendo a primeira o desenho esquemático e posteriormente a implementação do *layout*. Para a criação do esquemático é utilizado o *Composer Schematic*, seguindo-se a simulação esquemática e consequente extracção da *netlist* do circuito.

Caso os resultados obtidos se encontrem dentro dos parâmetros pretendidos, inicia-se através da ferramenta *Virtuoso Layout Editor*, o desenho do *layout*, que estabelece que todas as células criadas têm exactamente a mesma altura, variando apenas a largura segundo a complexidade da porta.

Depois do *layout* desenhado, a sua verificação é executada pelo DRC - *Design Rule Check*, que se certifica se as regras impostas pela tecnologia são cumpridas). Caso essas regras sejam verificadas, é feita uma extracção de capacidades/resistências parasitas usando o módulo RCX (*Resistance/Capacitance and Indutance Extraction*) por fim é feita uma verificação entre o *layout* e o *schematic* (LVS - *Layout Vs. Schematic*) para validar a correspondência entre as vistas.

Finalmente, já com o *layout* extraído, este é usado nos trabalhos de simulação incluindo neste caso já os elementos parasitas do desenho.

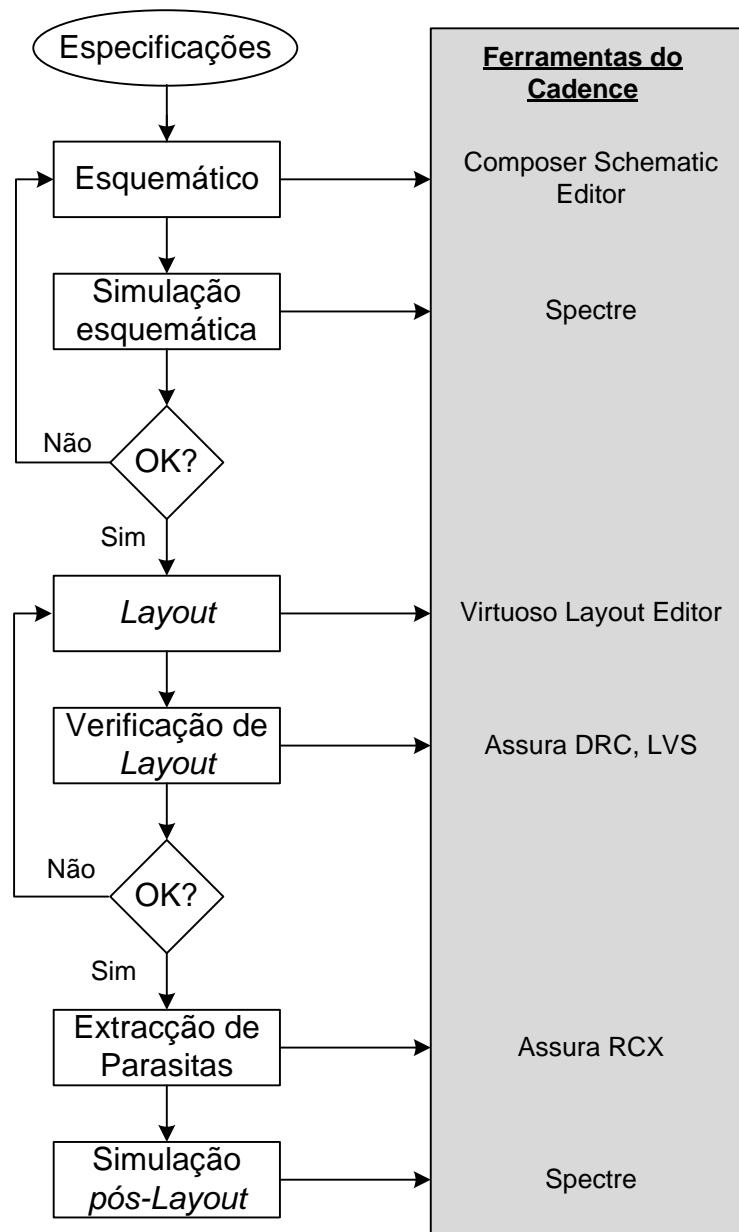


Figura 4.1 - Histograma sobre o funcionamento do Cadence como ferramenta de desenvolvimento

De ressaltar que inicialmente a verificação do *layout* era efectuada usando o *Assura*, mas problemas na extracção dos transístores obrigou a uma mudança da ferramenta de verificação para o *Diva*. De ressaltar que tanto o *Diva* como o *Assura* são ferramentas incluídas no *Cadence DF II*.

4.3 Conjuntos de Teste

O início deste trabalho coincidiu com o desenho de *layouts* de um amplo conjunto de portas lógicas CMOS estáticas. Nesta dissertação pretende-se estudar o impacto do *scaling* no desempenho de portas lógicas (inversores, NAND's e NOR's) de várias dimensões. Assim, neste estudo, foram usadas quatro dimensões diferentes para cada porta 1X, 2X, 4X, 8X.

Para se ter uma base de comparação entre as diversas portas usadas, começou por se projectar o inversor de tamanho mínimo (1X), aumentando posteriormente a largura dos PMOS e NMOS para 2X, 4X e 8X, mantendo constante a razão entre as dimensões dos transístores. Com base nas dimensões dos inversores de referência, criaram-se depois portas NAND2, NAND3, NOR2, NOR3.

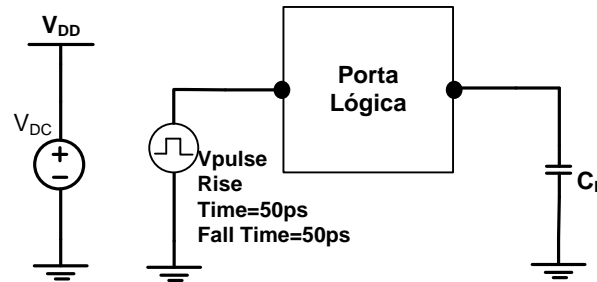


Figura 4.2– Ambiente de simulação com carga capacitiva

No primeiro teste analisou-se o comportamento das várias portas lógicas tendo como carga um condensador linear como representado na figura 4.2.

Este teste foi feito com o intuito de se obterem valores de referência a fim de se ter uma primeira ideia sobre o comportamento de cada porta e uma boa base de comparação sobre as simulações seguintes.

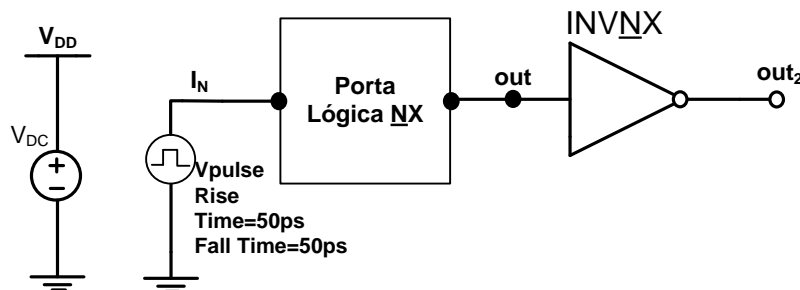


Figura 4.3 - Ambiente de simulação usando um inversor como carga

A figura 4.3 mostra o ambiente de teste utilizado na segunda experiência. Trata-se do esquemático com uma porta lógica a fazer o drive e um inversor a fazer de carga. Nesta simulação variou-se o tamanho do inversor de carga (1X, 2X, 4X, 8X) para cada tamanho da porta de drive.

Embora não fosse objectivo inicial, o principal caminho seguido na elaboração desta tese foi a compreensão e o estudo do efeito de *Miller*, nomeadamente o impacto deste na porta driver, bem como o seu efeito na porta de carga. Neste ponto foi usado um inversor ideal como driver (sem capacidades parasitas), idêntico aos inversões reais, de maneira a comparar com os valores obtidos na segunda experiência.

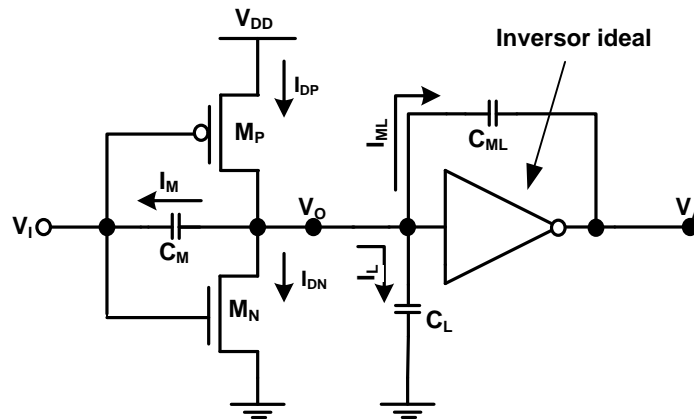


Figura 4.4 - Ambiente de simulação com um inversor ideal como carga

A figura 4.4 mostra o ambiente de simulação usado para a compreensão do efeito de *Miller*. Numa primeira instância o C_{ML} foi igualado a zero e C_L calculado manualmente fazendo uma média entre os valores para as duas regiões de funcionamento através de 4.1:

$$C_L = \frac{C_{gsP_{linear}} + C_{gsP_{saturado}} + C_{gsN_{linear}} + C_{gsN_{saturado}}}{2} \quad (4.1)$$

para saber a resposta de um inversor sem capacidades de *Miller* a fazer de carga.

Posteriormente estimou-se o valor de C_{ML} também manualmente usando 4.2

$$C_{ML} = \frac{C_{gdP_{linear}} + C_{gdN_{linear}}}{2} \quad (4.2)$$

para obter a resposta de uma porta lógica quando tem como carga um inversor real.

O esquema da figura 4.4 (com $C_{ML} = 0$) foi usado também para estudar o efeito da variação do VM na resposta das várias portas lógicas.

Por fim o esquema da figura 4.5 representa a última experiência efectuada, em que se usou um inversor real a fazer o drive de uma porta lógica real (NAND2 e NOR2) como cargas, variando a entrada à qual se liga o inversor.

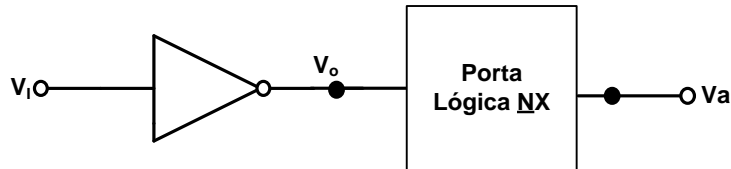


Figura 4.5 – Ambiente de simulação tendo como carga uma porta lógica

4.4 Resultados ao Nível do Layout

O trabalho de simulação teve por base o impacto que as várias cargas provocam no desequilíbrio dos tempos de propagação das portas driver. Foram testados inicialmente condensadores lineares como cargas, passando posteriormente para circuitos cujas cargas eram inversores. Testou-se também a influência que a capacidade de *Miller* e a variação de *threshold* das portas causam no desequilíbrio. Por fim e a título de curiosidade testaram-se portas lógicas mais complexas como cargas variando a entrada á qual era ligada a saída da porta *driver*.

4.4.1 Caracterização de um Inversor Estático

O parâmetro δ , indicado na equação 4.1, descreve o desequilíbrio que existe entre os tempos de propagação (t_{pHL} , t_{pLH}), e com ele pretende-se fazer uma comparação entre as várias portas lógicas desenhadas, por forma a facilitar a interpretação dos resultados.

$$\delta = 2 \frac{t_{pHL} - t_{pLH}}{t_{pHL} + t_{pLH}} \times 100(\%) \quad (4.1)$$

O dimensionamento dos transístores projectados foi através do ajuste dos parâmetros de desenho W_p e W_n , de forma a compensar a razão entre as mobilidades dos portadores do NMOS e PMOS isto é:

$$\beta = \frac{\mu_n}{\mu_p} = \frac{W_p}{W_n} = 3.25 \quad (4.2)$$

O valor 3.25 foi escolhido de forma a ter um bom equilíbrio entre os tempos de propagação, sendo assim considerado um valor empírico da tecnologia.

O inversor 1X foi construído utilizando um transistor NMOS com dimensões mínimas ($W_{\min} = 0.4\mu\text{m}$), enquanto ao PMOS foi aplicada a razão de desenho. Com base neste inversor, foram construídos outros três de dimensões 2X, 4X e 8X. O inversor 2X refere-se a um inversor com largura duas vezes superior à largura mínima.

O comprimento mínimo de canal foi mantido para todos os tamanhos das portas. Estas dimensões foram escolhidas uma vez que um inversor com o dobro da largura dos seus transístores manifesta uma condutância eléctrica duas vezes superior aproximadamente.

4.4.2 Estudo do Comportamento das Portas Lógicas com uma Carga Linear

O gráfico da figura 4.7, obtido usando o ambiente de simulação da figura 4.2, relaciona o desequilíbrio entre os tempos de propagação δ que as várias cargas capacitivas (variação entre 100fF e 10pF) provocam nas portas lógicas.

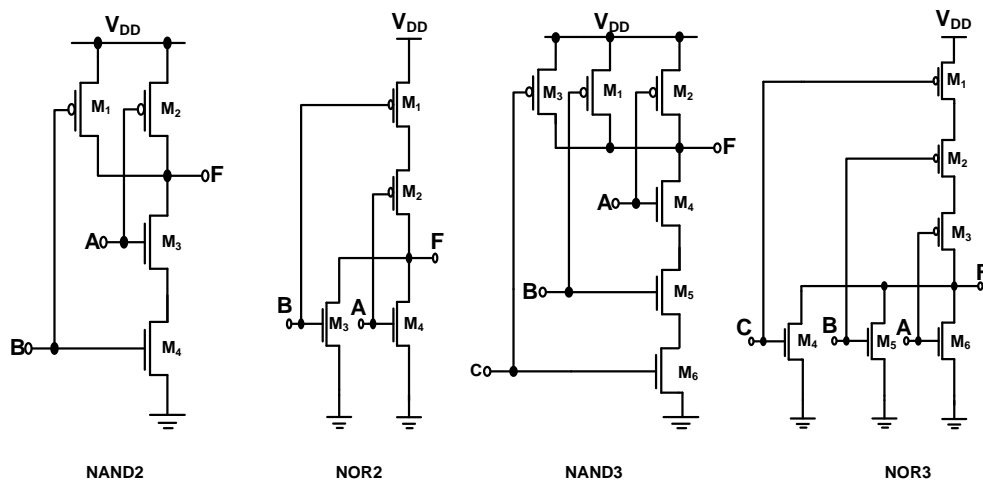


Figura 4.6 - Esquema eléctrico de uma NAND2, NOR2, NAND3 e NOR3

Estes resultados foram obtidos usando o pior caso para os tempos de propagação que para uma porta NAND2 ocorre quando $A=1$ e B transita de '0' para '1' (ver figura 4.6) e para uma porta NOR2 sucede se $A=0$ e B transitar de '1' para '0'. O pior tempo de propagação acontece para a NAND3 e para a NOR3 quando $A=B=1$ e C muda o estado de '0' para '1', $A=B=0$ e C a transitar de '1' para '0' respectivamente.

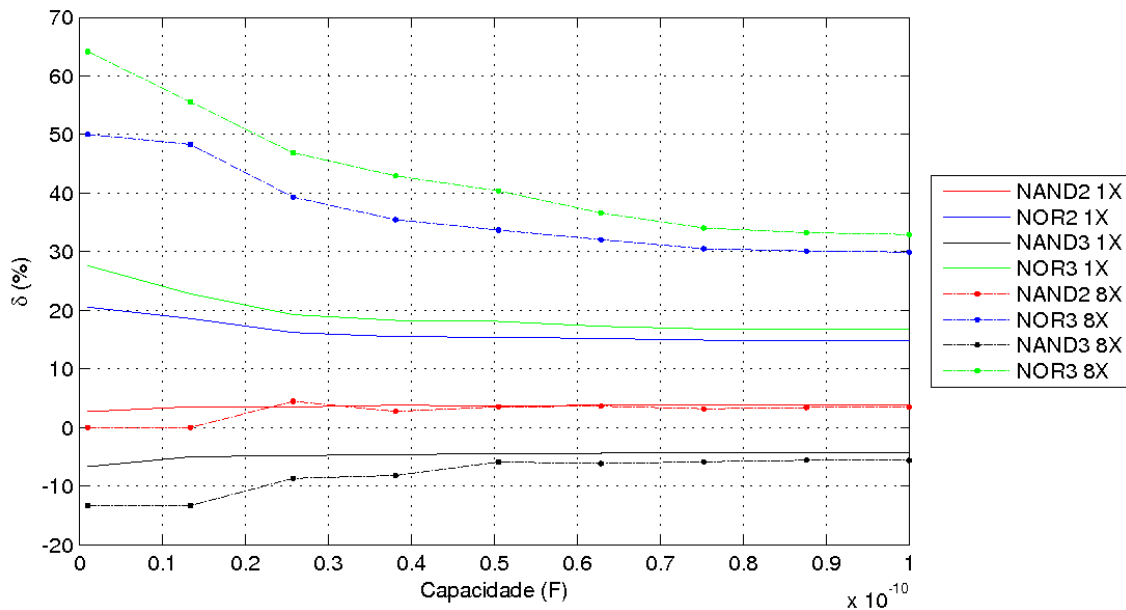


Figura 4.7 - Resposta das várias portas lógicas à variação de cargas capacitivas lineares

Como era expectável, uma vez que as cargas usadas são lineares, nas portas NAND o desequilíbrio é muito pequeno próximo de 5%, e quase constante para ambas as portas de 1X e 8X.

Já nas portas NOR, a divergência entre os tempos é enorme. Destes resultados podem tirar-se três conclusões: tal como já havia sido referido anteriormente, a série de PMOS causa grandes atrasos nas portas lógicas devido à mobilidade reduzida dos portadores de carga dos transístores PMOS; a segunda conclusão é a maior influência que, como explicado no capítulo 3, os transístores PMOS têm na capacidade de *Miller* uma vez que o seu tamanho é maior em relação aos NMOS, por fim a última conclusão a retirar é que valores elevados de δ , mostram que, neste caso, o valor de β a usar para o dimensionamento não devia ser 3.25.

Foi visível também que à medida que o tamanho das portas aumenta, os tempos de propagação diminuíram de forma quase constante, Anexo 1, já que as portas passam a ter mais capacidade de conduzir corrente.

4.4.3 Comportamento das Portas Lógicas Tendo um Inversor como Carga

Substituindo-se então a carga capacitiva por inversores reais (figura 4.3) de 4 tamanhos diferentes (1X, 2X, 4X e 8X) obteve-se o resultado ilustrado pela figura 4.8.

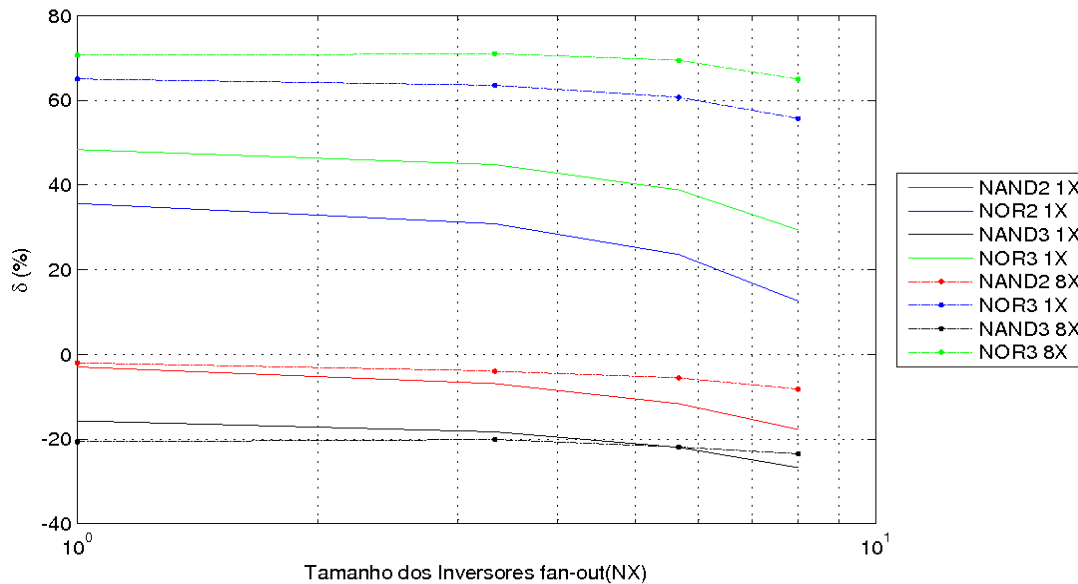


Figura 4.8 – Resposta das portas lógicas a um inversor de carga

Neste caso os desequilíbrios, quando comparados com os resultados da figura 4.7, são enormes e além disso os tempos de propagação (Anexo 1) não aumentam de forma proporcional ao aumento da dimensão da carga.

Uma vez que o estímulo de entrada tem um tempo de subida relativamente grande, 50ps, o efeito da capacidade de *Miller* da porta *driver* é relativamente pequeno podendo assim ser ignorado. Por outro lado, como estudado no capítulo 3 a capacidade de *Miller* no inversor de carga pode ter influência no comportamento dos circuitos, uma vez que os valores de δ não são lineares.

A capacidade total de carga (C_{LT}) vai ser influenciada por três valores: as capacidades parasitas do inversor *driver*, a capacidade de *Miller* da porta *driver* (que neste caso pode ser ignorada) e a capacidade de *Miller* do inversor de carga.

Como é visível também no gráfico 4.8 o desequilíbrio entre os tempos tem uma tendência decrescente e não linear: o tp_{HL} aumenta de forma moderada, em relação ao aumento do inversor carga.

Este aumento pode ser explicado com um aumento da carga, para o mesmo tamanho de porta driver, demorando assim mais tempo a descarregar o C_{IT} . Por outro lado o tp_{LH} , ou seja o tempo que o condensador demora a carregar, aumenta de forma mais acentuada. Este facto pode ser explicado pela presença de uma capacidade normalmente ignorada, a capacidade de *Miller*.

A capacidade de *Miller* carrega ao mesmo tempo que C_L , retirando-lhe assim corrente, fazendo aumentar o tp_{LH} . Além disso, como o C_{GD} do PMOS é bastante superior o C_{GD} do NMOS a capacidade de *Miller* faz-se sentir mais na transição ascendente.

Para confirmar estes resultados usou-se o ambiente de teste descrito pela figura 4.4 com C_L calculado usando a equação 4.1 e com $C_{ML} = 0$, ou seja sem existência de capacidade de *Miller*. O trabalho de simulação conduziu ao gráfico de resultados da figura 4.9.

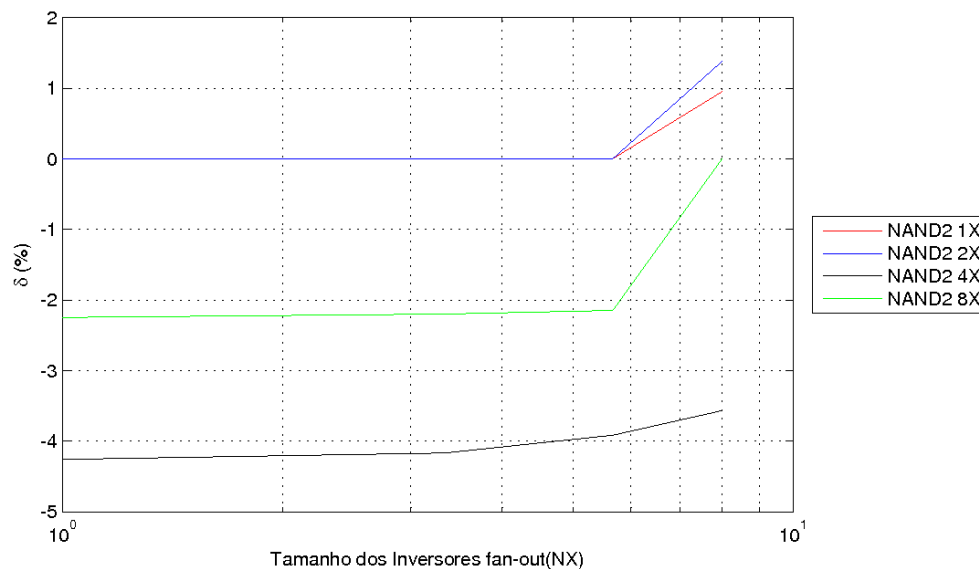


Figura 4.9 - Desequilíbrio da NAND2 em função de do tamanho do inversor ideal

Tal como no gráfico da figura 4.8, o δ das portas NAND2 é mínimo, cerca de 5% no pior caso, portanto estamos perante de um caso quase ideal. Para validar o efeito de

Miller observado na figura 4.8, voltou a simular-se o ambiente de teste 4.4, mas desta vez com um valor de C_{ML} atribuído usando a expressão 4.2. A figura 4.10 os resultados obtidos:

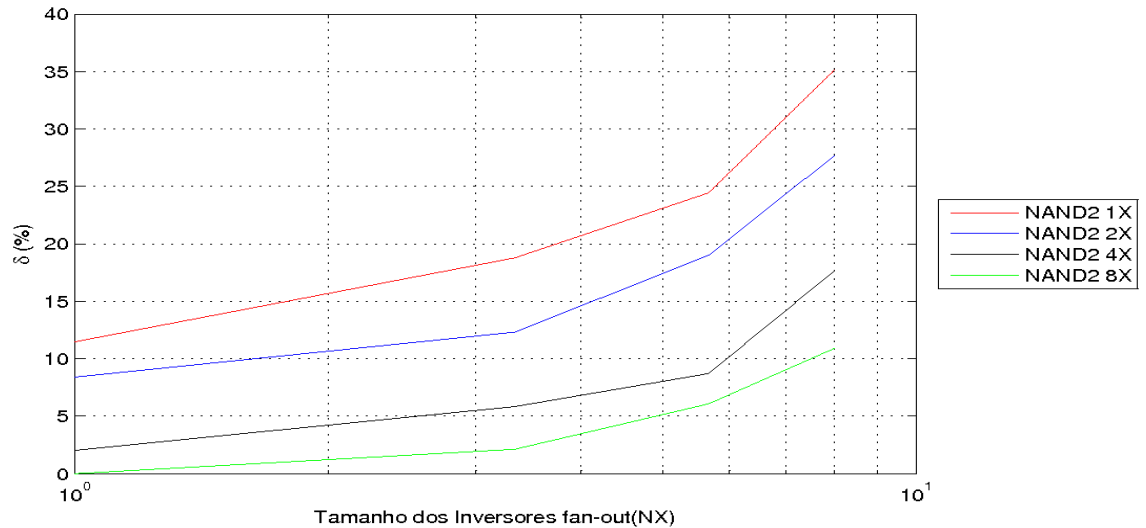


Figura 4.10 - Desequilíbrio em função do tamanho dos inversores ideais usando capacidade de Miller

A figura 4.10 demonstra que realmente o efeito de *Miller* no inversor de carga faz-se sentir nos desequilíbrios dos tempos de propagação. Quando se compara o δ da figura 4.10 com o δ da figura 4.8 é notório que são próximos, as diferenças existentes entre os valores do desequilíbrio deve-se ao facto de as capacidades terem sido calculadas de uma forma não muito precisa.

A orientação ascendente da curva quando se aumenta o tamanho do inversor de carga também é uma diferença visível. Devido a isso fez-se um ensaio para confirmar se era por culpa da influência do VM que a curva mudou de orientação. Voltou a realizar-se a experiência anterior tendo como porta driver uma NAND2_8X mudando o V_M da carga para $V_{M1} = V_M - V_M \times 20\%$ e $V_{M2} = V_M + V_M \times 20\%$, obtendo-se a figura 4.11.

Este gráfico mostra uma simetria em relação ao eixo das abcissas, quando se varia V_M . Com este resultado conclui-se que é possível compensar o efeito que a capacidade de *Miller* causa nos tempos de propagação e no seu equilíbrio, ajustando o valor de V_M da porta lógica de carga.

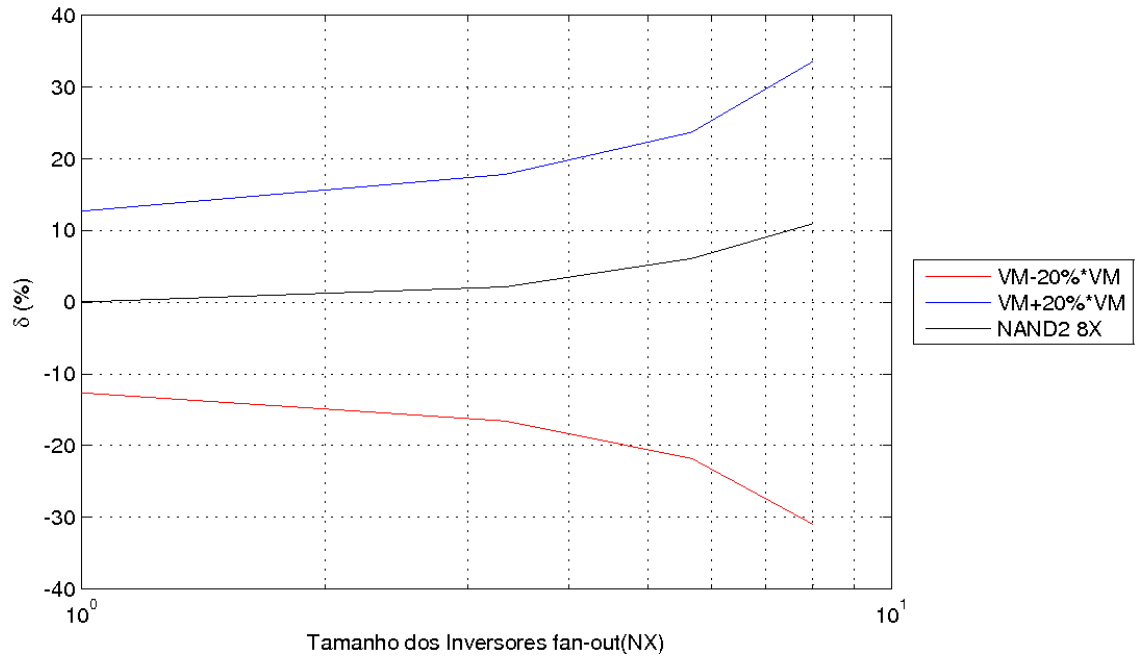


Figura 4.11 - Desequilíbrio em função do tamanho dos inversores para dois valores de VM diferentes tendo como driver um NAND2_8X

Por fim realizou-se, e a título de curiosidade realizou-se uma experiência usando o ambiente de simulação da figura 4.5. Os resultados obtidos foram os representados na figura 4.12.

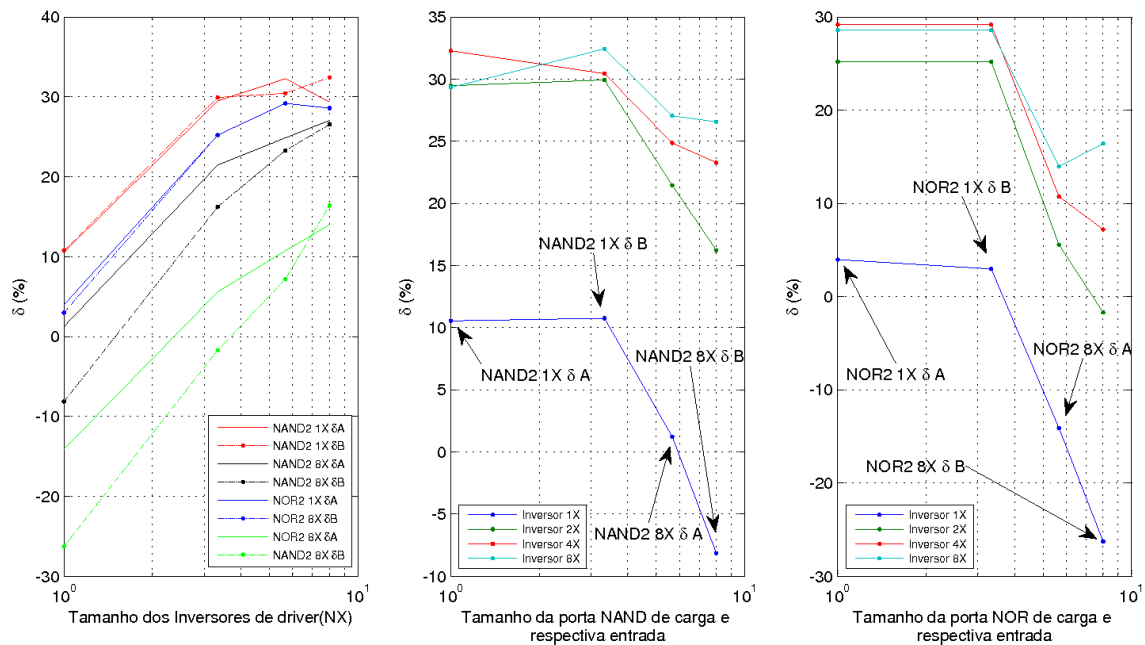


Figura 4.12 - Desequilíbrio em função do tamanho da porta driver (esquerda), Desequilíbrio em função de uma NAND como carga (meio), Desequilíbrio em função de uma NOR como carga (direita)

No gráfico é testado o comportamento do inversor real tendo como carga uma porta lógica mais complexa (NAND2, NOR2) de diferentes tamanhos e variando a entrada à qual se liga a saída do inversor.

Numa análise muito sucinta, é perceptível no gráfico da esquerda a enorme variação que os vários inversores de driver têm quanto à saída têm portas lógicas mais complexas. Por outro lado no gráfico do meio e da direita, é visível que não existe muita diferença de equilíbrios entre as portas de tamanho 1X mesmo que se varie a entrada à qual é ligada a saída do inversor (gráfico do meio e gráfico da direita) mas com dimensões maiores nas portas de carga, tornam a variação de desequilíbrios, quando se varia a entrada onde se liga o inversor driver, enorme

4.5 Resultados ao Nível da Simulação Esquemática

No desenvolvimento destas experiências foram usados dois esquemas de simulação, representados pelas figuras 4.2 e 4.3. A tecnologia usada foi novamente a AMS 350nm e o inversor de tamanho mínimo utilizado tinha um $W_n=0.4\mu m$ (dimensão mínima da tecnologia), $W_p=1\mu m$ e um $L=350nm$. Com a finalidade de se evitarem erros da tecnologia decidiu não se usar o inversor de tamanho mínimo como referência optando-se por um inversor 2X.

A figura 4.13 mostra a capacidade de entrada equivalente para um inversor com uma razão de desenho fixa e de valor 2.5, quando W_p varia de $2\mu m$ para $8\mu m$. Esta capacidade equivalente foi calculada para transições HL e LH e por fim foi calculada a média do set-up de referência (figura 4.2).

A figura 4.13 apresenta uma relação praticamente linear quando a carga é um condensador linear ou um inversor. Assim foi possível fazer as experiências seguintes, estabelecendo uma relação directa entre C_{eq} e W_p/W_n .

A relação é dada pela equação 4.3:

$$\frac{C_{eq}}{W_p} \simeq 1.95 (Ff/\mu m) \quad (4.3)$$

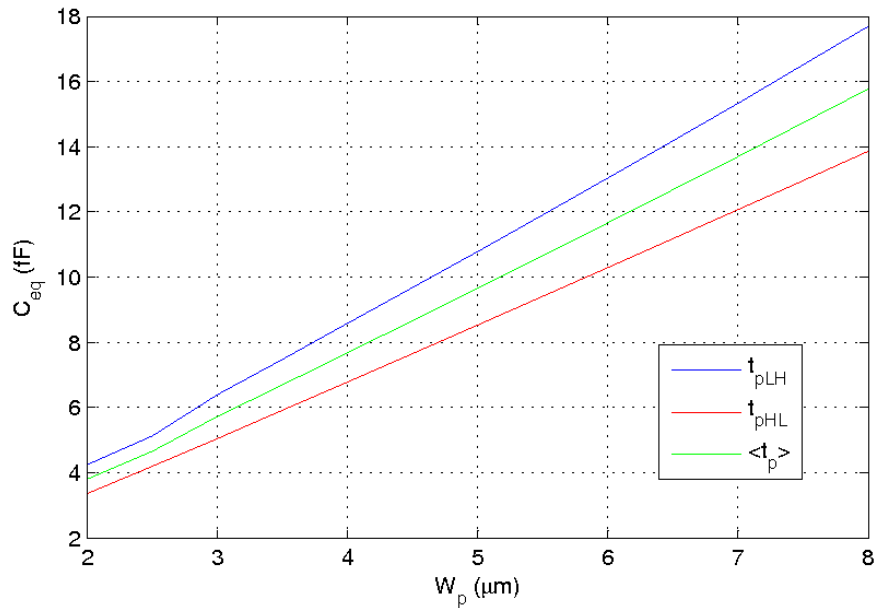


Figura 4.13 – Capacidade de Entrada equivalente

Utilizando novamente a comparação entre os mesmos dois esquemas de simulação para se ter uma percepção da evolução da capacidade de Miller com o tempo de entrada, fez-se variar o tempo de entrada do circuito entre 1ps e 500ps e obteve-se a figura 4.14. A comparação foi feita em termos do desequilíbrio em relação t_{rin} (tempo de subida do sinal de entrada). As curvas visíveis correspondem ao valor de três capacidades de carga lineares (gráfico da esquerda) que correspondem as curvas três inversores de carga, 2X, 4X e 8X.

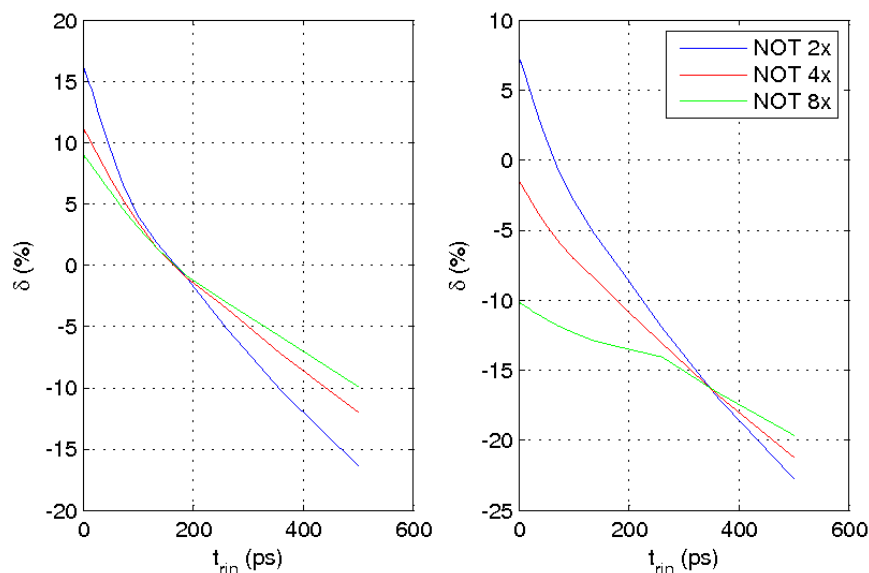


Figura 4.14 – Dependência entre os tempos de propagação e o tempo de subida (esquerda – esquema de simulação da figura 4.2, direita esquema de simulação da figura 4.3)

É notório o desequilíbrio causado pela variação de t_{rin} , mas o impacto da variação das duas condições de carga também não pode ser ignorado.

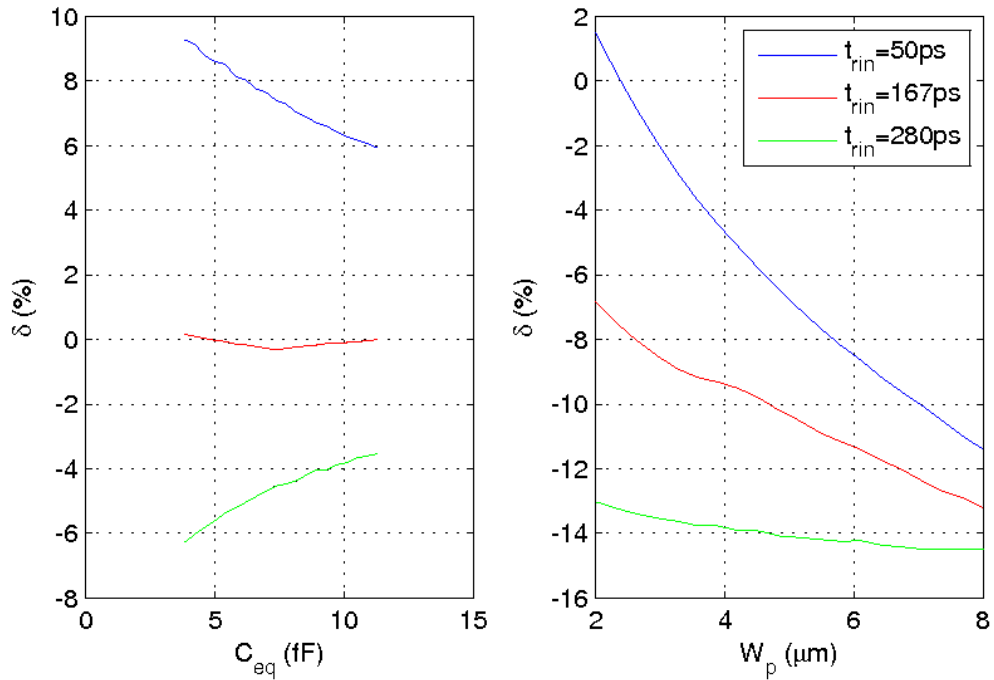


Figura 4.15 – Desequilíbrio em relação às condições de carga (esquerda – esquema de simulação da figura 4.2, direita esquema de simulação da figura 4.3)

Considerando as condições de carga como a variável, testou-se os dois esquemas de simulação (figura 4.2 e figura 4.3) para três diferentes tempos de subida (tendo a mesma capacidade equivalente do teste anterior) o que deu origem à figura 4.15. Com o esquema da figura 4.2 o desequilíbrio para um tempo de subida de 167ps, é aproximadamente 0%, e para o tempos de subida 50ps e 280ps, há algum desequilíbrio que diminui quando a capacidade de carga equivalente aumenta.

Já para o gráfico da direita a realidade é outra, para os três tempos de subida, o desequilíbrio é notório e o circuito deixa de responder de forma igual quando comparado com o circuito que tinha como carga uma capacidade linear.

Para o último teste utilizou-se apenas o esquema da figura 4.3 e foram assumidas razões de desenho diferentes para o inversor de *driver* (r_D) e para o inversor de carga (r_L).

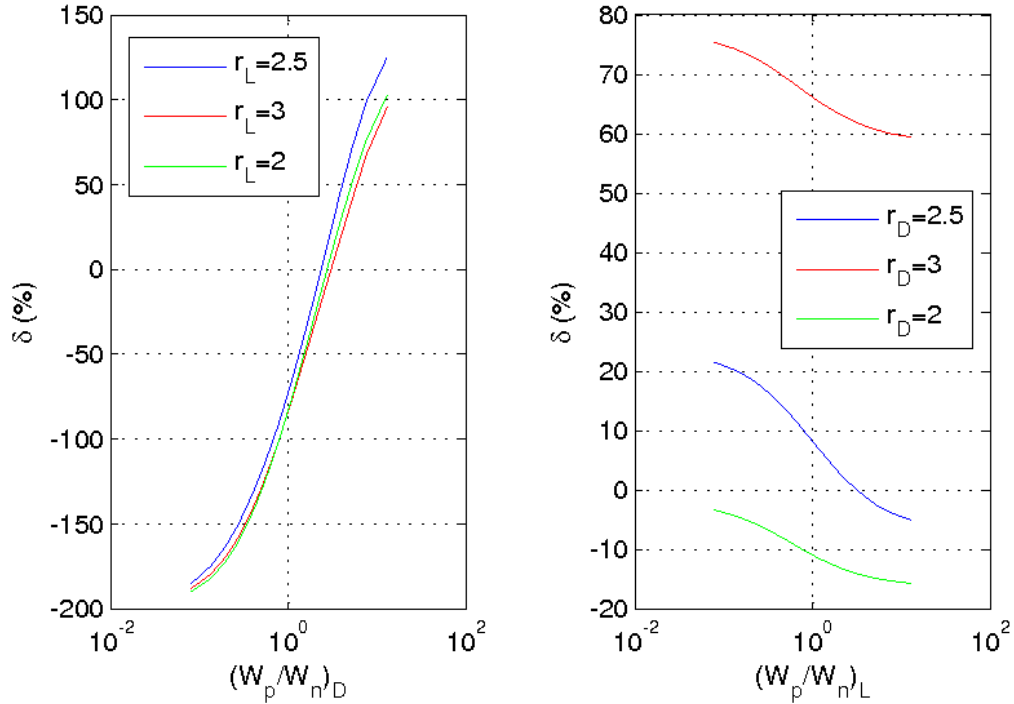


Figura 4.16 – Desequilíbrio em função das razões de desenho (esquerda – inversor de driver, direita – inversor de carga)

A figura 3.16 é constituída por dois gráficos, o gráfico da esquerda representa uma experiência em que o inversor de carga foi fixo em três razões de desenho (2, 2.5 e 3) variando para cada caso a razão de desenho do inversor de *driver*. O gráfico da direita é o resultado uma experiência em que se mantiveram constantes três razões de desenho do inversor de *driver* e fez-se variar a razão de desenho do inversor de carga para cada uma delas.

Como era expectável, a variação da razão de desenho no inversor de *driver* trouxe um desequilíbrio enorme entre os tempos de propagação, por outro lado a variação da razão de desenho no inversor de carga também desequilibrou muito os tempos de propagação o que revela a importância e o cuidado que é necessário ter quando se coloca um inversor como carga. De notar que quando se coloca uma porta lógica mais complexa como carga, a complexidade destes resultados aumenta uma vez que há mais capacidades e mais combinações possíveis para a activar.

4.6 Sumário

Neste capítulo foi descrito o estudo do comportamento dinâmico de várias portas lógicas utilizadas. Este baseou-se numa simulação esquemática numa primeira instância seguido do desenho do layout. Foram descritas também as várias estratégias seguidas ao longo do trabalho e igualmente foram apresentados os resultados e os respectivos comentários.

Os vários resultados obtidos com o trabalho de simulação, foram obtidos ao nível do layout e demonstram que embora não existam grandes referências na literatura quanto ao efeito de *Miller* provocado pela carga, este não pode ser desprezada, uma vez que contribui para as divergências causadas no equilíbrio das várias portas CMOS. Foi também estudada uma solução para combater o efeito da capacidade de *Miller*.

Em suma, foram obtidos resultados conclusivos relativamente ao efeito de *Miller* e o efeito deste no desenho de circuitos lógicos, bem como a influência que a razão de desenho tem nos circuitos.

Capítulo 5 – Conclusões

Nesta dissertação foram discutidas técnicas/efeitos provocadas pelo *scaling* no desenho de circuitos lógicos. A evolução das tecnologias CMOS foi acompanhada por alguns efeitos que obrigaram o estudo e a criação de novos modelos analíticos e de desenho.

O estudo feito ao longo deste trabalho assentou na perspectiva de criação e estudo de um modelo matemático que justificasse o efeito causado pela variação de VM da porta lógica e da capacidade de *Miller*. Uma variação na razão de mobilidades de uma porta lógica influencia o VM da porta e com isto, o tempos de propagação também variam. Com este estudo é possível perceber como alguns factores por vezes ignorados podem interferir no desenho de circuitos digitais.

Dissecando este trabalho, verifica-se que a parte teórica estudada nos primeiros capítulos justifica os resultados práticos obtidos. Começou por se analisar os efeitos secundários que se notam principalmente nas tecnologias de canal curto. Posteriormente estudaram-se os alicerces em que a tecnologia CMOS está imbuída bem como os conceitos base fundamentais sobre o comportamento dos circuitos de electrónica digital. Por fim foi realizado um estudo sobre a capacidade de *Miller* e o VM e a influência destes sobre os atrasos das portas lógicas.

O trabalho de simulação realizado bem como os resultados obtidos ofereceram um conjunto de resultados importantes para a compreensão dos efeitos inerentes à construção física da porta.

De destacar que este trabalho abrangeu o desenho físico das várias portas, isto é, desenho de *layout*. É notório que existe alguma divergência quando se trabalha ao nível do esquemático apenas, quando comparado com o trabalho ao nível da parte física, sendo neste caso importante boas práticas de desenho bem como coerência entre as várias portas, uma vez que o circuito desenhado causa o aparecimento de capacidades parasitas que tem importância na simulação pós-*layout*.

5.1 Linhas de investigação futuras

Como se conclui deste estudo, a capacidade de *Miller* bem como um mau ajuste do VM das portas causa grandes desequilíbrios entre os tempos de propagação.

Num estudo futuro seria curioso uma análise que permitisse calcular um factor de ajuste para o VM de cada porta lógica, isto é, desenhar a porta lógica com o mínimo de desequilíbrio possível, devido a estes efeitos, entre os tempos de propagação.

Seria também interessante fazer um estudo mais exaustivo e baseado em modelos teóricos para o comportamento das várias portas tendo como carga uma porta mais complexa como carga e ver o comportamento do circuito quando a entrada que é “atacada” na carga varia.

Era interessante também uma repetição do estudo desta tese mas utilizando outras tecnologias e repetir as experiências feitas ao nível do esquemático, mas desta vez realizadas ao nível do layout.

Referências

1. Frank M. Wanlass and Chih-Tang Sah, "Nanowatt Logic Using Field-Effect Metal-Oxide Semiconductor Triodes", International Solid-State Circuits Conference, pp. 32–33, February 1963.
2. Harold Shichman and David A. Hodges, "Modeling and Simulation of Insulated-Gate-Field-Effect Transistor Switching Circuits", IEEE Journal of Solid-State Circuits, Vol.SC-3, No. 3, pp. 285–288, September 1968.
3. Yannis P. Tsividis, Operation and Modeling of the MOS Transistor, International Editions, McGraw-Hill, 1988.
4. J. E. Meyer, "MOS models and circuit simulation," RCA Rev., vol.32, pp. 42-63, 1971.
5. L. W. Nagel, "SPICE2: A computer program to simulate semiconductor circuits," ERL-M520, Electronics Res. Lab., Univ. Calif., Berkeley, 1975.
6. D. E. Ward and R. W. Dutton, "A charge oriented model for MOS transistor capacitances," IEEE J. Solid-State Circuits, vol. SC-13, pp. 703-707, 1978
7. Kaushik Roy and Sharat C. Prasad, Low-Power CMOS VLSI Circuit Design, Wiley-Interscience, 2000.
8. J. E. Meyer, "MOS models and circuit simulation," RCA Rev., vol. 32, pp. 42-63, 1971.
9. L. W. Nagel, "SPICE2: A computer program to simulate semiconductor circuits," ERL-M520, Electronics Res. Lab., Univ. Calif., Berkeley, 1975.
10. D. E. Ward and R. W. Dutton, "A charge oriented model for MOS transistor capacitances," IEEE J. Solid-State Circuits, vol. SC-13, pp. 703-707, 1978
11. Kunz V.D., Uchino T., de Groot C.H., Ashburn P. et al.; Reduction of Parasitic Capacitance in Vertical MOSFETs by Spacer Oxidation; IEEE Trans Electron Devices 2003; 50(6); p.1487-1493
12. J. S. Kilby, "Miniaturized Electronic Circuits", U. S. Patent 3138743 February 6, 1959.

13. R. N. Noyce, "Semiconductor Device-and-Lead Structure", U. S. Patent 2981877, July 30, 1959.
14. G. E. Moore, "Cramming More Components Into Integrates Circuits", Electronics, pp.32-33, February 1963.
15. Robert H. Dennard, Fritz H. Gaensslen, L. Kuhn and H. N. Yu, "Design of micron MOS switching devices" IEDM Technical Digest, Vol. SC-3, No. 3, pp. 168-170, December 1972.
16. "www.Intel.com" pesquisado no mês de Setembro de 2009
17. J. Araújo, "Impacto do scaling da tecnologia CMOS no desenho de circuitos digitais," M.S. thesis, Universidade de Aveiro, Aveiro, Portugal, 2008
18. Jan M. Rabaey, Anantha Chandrakasan and Borivoje Nikolic, Digital Integrated Circuits: A Design Perspective, Second Edition, Internacional Edition, Prentice Hall
19. Neil H. E. Waste and Kamran Eshraghian, *Pinciples of CMOS VLSI Design: A System Perspective*, Second Edition, Addison-Wesley Publishing Company, 1993.
20. Risch L., Krautschneider W.H., Hofmann F., SchäferH., Aeugle T., Rösner W.; IEEE Trans. Electron Devices, vol 43, p1495 (1996)
21. Choi Y-K, King T-J, Hu C.; Nanoscale CMOS spacer FINFET for the terabit era; IEEE Electron Device Letters 2002; 23; p.25-27
22. Robert H. Dennard, Fritz H. Gaensslen, H. N. Yu, V. Leo Rideout, Ernest Bassous and Andre R. Leblanc, "design of ion-implanted MOSFET's with Very Small Physical Dimensions", IEEE Journal of Solid-State Circuits, Vol. SC – 9, No. 3, pp. 168-170, December 1972
23. Gopalan K., Introduction to Digital Microelectronic Circuits, McGraw-Hill Primis
24. Boris Andreev, Edward L. Tittlebaum e Eby G. Friedman, "Sizing CMOS Inverters with Miller Effect and Threshold Voltage Variations", Journal of Circuits and Computers, Vol. 15, No.3 (2006), pp.437-454, March 2006.
25. Sérgio Pires, M. Dionísio Rolo, Luís Nero Alves, Ernesto Martins, "A Study on the Propagation Times of Loaded CMOS Inverters", 2009, submetido para a conferencia Iscas 2010, www.iscas2010.org.
26. Phaeton Avouris, J. Appenzeller, R. Martel and S. Wind, "Carbon Nanotube Electronics", Proceedings of the IEEE, Vol.91, pp. 1772-1784, 2003.

Anexos 1

NAND2_1x → PMOS: $W = 1.3\mu\text{m}$ **NMOS:** $W = 0.8\mu\text{m}$

	$T_{pHL}(\text{ns})$	$T_{pLH}(\text{ns})$	$\delta (\%)$
Inversor 1x	0.079	0.083	-4.94
Inversor 2x	0.093	0.101	-8.25
Inversor 4x	0.123	0.140	-12.92
Inversor 8x	0.180	0.220	-20.00

NAND2_8x → PMOS: $W = 10.4\mu\text{m}$; **NMOS:** $W = 6.4\mu\text{m}$

	$T_{pHL}(\text{ns})$	$T_{pLH}(\text{ns})$	$\delta (\%)$
Inversor 1x	0.045	0.047	-4.38
Inversor 2x	0.047	0.049	-4.17
Inversor 4x	0.050	0.053	-5.83
Inversor 8x	0.057	0.062	-8.40

NOR2_1x → PMOS: $W = 2.6\mu\text{m}$; **NMOS:** $W = 0.4\mu\text{m}$

	$T_{pHL}(\text{ns})$	$T_{pLH}(\text{ns})$	$\delta (\%)$
Inversor 1x	0.162	0.113	35.64
Inversor 2x	0.176	0.129	30.82
Inversor 4x	0.204	0.161	23.56
Inversor 8x	0.262	0.231	12.58

NOR2_8x → **PMOS**: $W = 20.8 \mu\text{m}$; **NMOS**: $W = 3.2 \mu\text{m}$

	TpHL(ns)	TpLH(ns)	δ (%)
Inversor 1x	0.110	0.056	65.06
Inversor 2x	0.112	0.058	63.53
Inversor 4x	0.116	0.062	60.67
Inversor 8x	0.124	0.070	55.67

NAND3_1x → **PMOS**: $W = 1.3 \mu\text{m}$; **NMOS**: $W = 1.2 \mu\text{m}$

	TpHL(ns)	TpLH(ns)	δ (%)
Inversor 1x	0.105	0.125	-17.39
Inversor 2x	0.117	0.144	-20.69
Inversor 4x	0.145	0.183	-23.17
Inversor 8x	0.201	0.265	-27.47

NAND3_8x → **PMOS**: $W = 10.4 \mu\text{m}$; **NMOS**: $W = 9.6 \mu\text{m}$

	TpHL(ns)	TpLH(ns)	δ (%)
Inversor 1x	0.063	0.077	-20.00
Inversor 2x	0.065	0.080	-20.69
Inversor 4x	0.068	0.084	-21.05
Inversor 8x	0.074	0.094	-23.81

NOR3_1x → PMOS: $W = 3.9 \mu\text{m}$; NMOS: $W = 0.4 \mu\text{m}$

	TpHL(ns)	TpLH(ns)	δ (%)
Inversor 1x	0.280	0.171	48.34
Inversor 2x	0.295	0.187	44.81
Inversor 4x	0.326	0.220	38.83
Inversor 8x	0.387	0.288	29.33

NOR3_8x → PMOS: $W = 31.2 \mu\text{m}$; NMOS: $W = 3.2 \mu\text{m}$

	TpHL(ns)	TpLH(ns)	δ (%)
Inversor 1x	0.090	0.043	70.68
Inversor 2x	0.092	0.044	70.59
Inversor 4x	0.097	0.047	69.44
Inversor 8x	0.106	0.054	65.00